Towards structure elucidation of the Proline Reductase Complex from *Clostridioides difficile*

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

Dissertation

von Maurice Günther Diwo
aus Bad Homburg vor der Höhe

1. Referent:	Professor Dr. Wulf Blankenfeldt
2. Referent:	Professor Dr. Dieter Jahn
eingereicht am:	10.03.2021
mündliche Prüfung (Disputation) am:	28.05.2021

Druckjahr 2021

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit,
in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Maurice Diwo, Wiebke Michel, Philipp Aurass, Katja Kuhle-Keindorf, Jan Pippel, Joern Krausze, Christina Lang, Wulf Blankenfeldt, Antje Flieger. NAD(H)-mediated tetramerization controls the activity of *Legionella pneumophila* phospholipase PlaB. PNAS June 8, 2021, 118 (23) e2017046118. DOI: 10.1101/2020.09.01.246603

Tagungsbeiträge

Maurice Diwo, Wiebke Michel, Philipp Aurass, Katja Kuhle-Keindorf, Jan Pippel, Joern Krausze, Christina Lang, Wulf Blankenfeldt, Antje Flieger. Phospholipases from *Legionella pneumophila.* Heart of European Biocrystallography Meeting 21, Quedlinburg, 20.09.2018.

Maurice Diwo, Wiebke Michel, Philipp Aurass, Katja Kuhle-Keindorf, Jan Pippel, Joern Krausze, Christina Lang, Wulf Blankenfeldt, Antje Flieger. Structure and Function of the phospholipase PlaB from *Legionella pneumophila.* 5th Workshop of Structural Biology in the Helmholtz Association, 17.12.2019.

Posterbeiträge

Maurice Diwo, Julia Hofmann, Kim Rennhack, Mathias Müsken, José Borrero de Acuña, Meina Neumann-Schaal, Alexander Marchanka, Andreas Pich, Wulf Blankenfeldt. Structural and functional insights to Stickland fermentation in *Clostridium difficile* (Poster). Heart of European Biocrystallography Meeting 20, Jelenia Gora, 28.09.2017.

Maurice Diwo, Wiebke Michel, Philipp Aurass, Katja Kuhle-Keindorf, Jan Pippel, Joern Krausze, Christina Lang, Wulf Blankenfeldt, Antje Flieger. Crystal Structure of the phospholipase PlaB, a hemolytic virulence factor from *L. pneumophila.* Deutsche Gesellschaft für Kristallografie, Essen, 06.03.2018.

Maurice Diwo, Wiebke Michel, Philipp Aurass, Katja Kuhle-Keindorf, Jan Pippel, Joern Krausze, Christina Lang, Wulf Blankenfeldt, Antje Flieger. Towards structure elucidation of the Proline Reductase Complex from *Clostridioides difficile.* CLOSTPATH 11, Leiden, 08.08.2019.

I. INDEX

II. List of Abbreviations

AAD	antibiotic associated diarrhea
AC	affinity chromatography
Amp	ampicillin
BESSY	Berliner Elektronensynchotron
BME	β-mercaptoethanol
CD	circular dichroism
CDI	*Clostridioides difficile* infection
CDMM	*Clostridioides difficile* mineral medium
Cm	chloramphenicol
Ctf	contrast transfer function
DESY	Deutsches Elektronen Synchrotron
DLS	dynamic light Scattering
DMSO	dimethyl sulfoxid
DNA	desoxyribonucleic acid
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DSSO	disuccinimidyl sulfoxide
EDTA	ethylenediaminetetraacetic acid
EPU	E pluribus unum (lat. one made of many)
ESRF	european synchrotron radiation facility
GA	glutaraldehyde
gDNA	genomic DNA
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIC	hydrophobic interaction chromatography
HZI	Helmholtz-Centre for Infection Research
IEX	ion exchange chromatography
IR	infra-red
Kan	kanamycin
LB	lysogeny broth
M9MM	M9 mineral medium
MALS	multi angle light scattering
MOPS	3-(N-morpholino)-propanesulfonic acid
MR	molecular replacement
MS	mass spectrometry
MST	microscale thermophoresis
MW	molecular weight
MWCO	molecular weight cut-off
NAD	nicotineamidedinocleotide
NMR	nuclear magnetic resonance
NTA	nitrilotriacetic acid
o/n	over night
OD	optical density
PAGE	polyacrylamide gelelectrophoresis
PaLoc	pathogenicity locus

PCR	polymerase chain reaction
PEG	polyethylene glycol
PIPE	polymerase incomplete primer extension
PVDF	polyvinylidene fluoride
RKI	Robert Koch Institut
RPEX	Recombinant Protein Expression
RT	room temperature
SAD	single anomalous dispersion
SAXS	small angle x-ray scattering
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
SFPR	Structure and Function of Proteins
SLS	swiss light source
SNAD	thio-nicotinamidedinucleotide
SOB	super optimal broth
SOC	super optimal broth with catabolite repression
SS	secondary structure
TAE	TRIS Acetate EDTA buffer, pH 8
TB	terrific broth
Tcd	toxin of *Clostridioides difficile*
TEM	transmission electron microscopy
TEV	tobacco etch virus
TEV	tobacco etch virus
TFB	transformation buffer
Tm	melting temperature
TRIS	tris-(hydroxymethyl)-aminomethane
TSA	thermal shift assay
UV	ultra violet
WB	western blot
WT	wild type

III. List of Figures

IV. List of tables

V. Summary

The Gram positive and obligate anaerobe bacterium *Clostridioides difficile* can be found ubiquitously in our environment. Simultaneously, *C. difficile* infections (CDI) are the most common cause of hospital-acquired, antibiotic-induced diarrhea. CDI leads to severe diarrhea, colon inflammation and eventually to life-threatening colon perforation. Antibiotic resistance of *C. difficile* and increasingly poor response to treatment are associated with increasing numbers of cases in hospitals in industrialized countries. These circumstances justify the desperate search for innovative treatment options that do not rely on traditional broadband antibiotic dosages to alleviate CDI infection courses. Within the last years, *C. difficile* attracted increasing attention from multi-disciplinary research, creating a steady fundament for a better understanding. However, *C. difficile* still raises exciting questions that need to be answered urgently.

C. difficile gains its ability to overgrow and dominate compromised intestinal flora by several specialized metabolic pathways. One of such is known as the Stickland fermentation. Here, amino acids are used in a coupled reaction as source of carbon and energy. Oxidative reactions include deamination and decarboxylation of various amino acids under generation of ATP and NADH. Reductive reactions occur specifically and mainly on proline and glycine. Metabolic investigations showed that proline is the preferred substrate for NAD^+ regeneration and the proline reductase has been shown to play important roles in manifold metabolic regulations. An intact proline reductase has been proven to be a requirement for C. *difficile* growth and sporulation. Further, toxin production is regulated by PrdR, the regulator of the proline reductase operon. This σ54-dependent regulator constitutes a direct link of proline availability and therefore the metabolic state of the pathogen and induction of toxin production.

The proline reductase itself was described to consist of three subunits, PrdAβ, PrdAα and the selenocystein containging PrdB. The β and α subunits mature by a rare selenocysteinolysis from the PrdA-proprotein. The complex was confirmed to be soluble, but previous studies also showed an association to the bacterial membrane, presumably indicating interactions to membrane proteins. The single subunits have molecular weights (MW) of 45, 23 and 25 kDa, respectively, and were found to assemble into a large supercomplex of more than 800 kDa. However, due to its exclusiveness for *Clostridia* and prevailing poor accessibility of the proteins, detailed molecular understanding for structure

and function of the proline reductase supercomplex remains elusive. Comprehensive and detailed knowledge of the assembly, the architecture and molecular functionality can potentially contribute to identification of alternative approaches for CDI treatment.

The most important prerequisite for biochemical and structural characterization is a valid and robust protein isolation procedure. Existing purification protocols for the proline reductase supercomplex were revised and refined to obtain a sample of the highest possible quality, in order to meet the requirements for structural elucidation by negative stain TEM. The proline reductase supercomplex was isolated directly from *C. difficile* to ensure correct folding and complex assembly. Biophysical investigations by SAXS, MALS, negative stain TEM, as well as sequence-based analyses allowed describing the proline reductase supercomplex as a metastable triacontameric complex of 930 kDa, exhibiting a large extent of intrinsic flexibility. Results from limited proteolysis suggested that an N-terminal stretch of PrdAβ is responsible for pentamerization of hexamers. The remaining parts of PrdAβ and PrdAα comprise an interaction scaffold for PrdB. Together, the hexamer $PrdA\beta_2\alpha_2PrdB_2$ constitutes a rather rigid catalytically active domain, which is congruously represented with five copies in the full supercomplex. For both, truncated but unmatured PrdA-proprotein and PrdB, recombinant isolation procedures were established. These strategies allowed generation of preliminary cryo-EM data that will likely lead to a high-resolution structure of the truncated, hexameric subcomplex in the near future.

Besides *C. difficiles*’s proline reductase, another focus was put on the phospholipases originating from *Legionella pneumophila*. This Gram-negative bacterium naturally occurs in cold fresh-water reservoirs, where it is able to replicate inside amoebae by modulation of host cell functions and vesicle trafficking. This ability arises from a highly complex secretome, a large number of modulators that are injected into the host cell or the *Legionella* containing vacuole (LCV). The same mechanisms enable *L. pneumophila* to infect human lung macrophages upon inhalation of contaminated aerosols, causing heavy pneumonia, described as the so-called Legionnaires disease.

Important examples of *L. pneumophila*’s secretome and dominant virulence factors are phospholipases. These enzymes attack host cell membranes and thereby alter membrane integrity and release toxic second messengers. In order to protect itself from noxious activity, the phospholipases are subjected to strict regulation mechanisms.

The phospholipase PlaA is secreted from *L. pneumophila* and activated in the LCV lumen by proteolytic maturation. A backbone cleavage is conducted by ProA, a secreted and rather unspecific protease, also acting as a virulence factor. Structural determination of pre-mature PlaA was achieved by X-ray crystallography. It unveiled that the lid domain,

framed and stabilized by a disulfide bond, shields the active site from the solvent. The proteolysis site was found at an exposed location in the lid domain. Hence, the maturation event is expected to increase lid flexibility. At the same time, the disulfide bond maintains the enzyme’s structural integrity despite backbone cleavage. This way, the active site is unshielded and the triggering of phospholipase activity can be explained.

Another sophisticated principle of enzymatic regulation for protection from self-inflicted lysis was revealed by characterization of the highly potent virulence factor PlaB. The tetrameric structure obtained from X-ray crystallography allowed the generation of structure-guided mutants that were studied by a combination of sub-cellular localization experiments and activity assays. Together, these data suggest that PlaB activity is inhibited intracellularly by a novel mechanism of NAD(H)-mediated dimer-of-dimer formation. Upon NAD(H) detachment, the dimer/dimer interface becomes unstable. The monomer/monomer interface within the dimer, however, is built from an almost 30 amino acid long stretch at the C-terminus of PlaB. This hook-like structure establishes a stable interaction between both monomers and is essential for enzymatic activity due to hydrolase-fold complementation. An unknown export mechanism transfers the enzyme to the bacterial surface, where non-canonical structural elements facilitate membrane association of PlaB dimers. These dimers represent the active form of PlaB that exhibit their lytic potential.

1 Introduction

1.1 Topic I: The proline reductase from *Clostridioides difficile*

The proline reductase (PR) of *Clostridioides difficile* is a central metabolic enzyme with massive influence on many functions. Studies have shown necessity for growth, clear participation in toxin regulation and tight connection to sporulation. Thus, all three major topics in *C. difficile* research are touched by this entity, highlighting the importance for the organism.

1.1.1 *Clostridioides difficile*, an overview on an exceptional bacterium

Clostridioides difficile, formerly known as *Clostridium difficile*, is an obligate anaerobe, Gram-positive, rod-shaped and spore forming bacterium (for microscopic image, see figure 1.1). It was initially identified from the intestinal flora of newborns as *Bacillus difficilis. The designation was* inspired by troublesome cultivation behavior under anaerobic conditions (Hall et al., 1935). For historical reasons, the class of *Clostridia* is divided in 14 clusters, covering highly divergent anaerobes. For a long time, 16 sRNA ribotyping provided a basis for classification of *Clostridia* (Gurtler, 1993). However, strains classified to be closely related did not necessarily show the same phenotypes and exhibited large differences in pathogenicity. At the same time, highly divergent strains showed similar results for ribotyping (Sirard et al., 2011). Consequently, a set of miscellaneous methods for classification of *Clostridia* were developed. As sequencing capacities increased, the most comprehensive classification by whole genome analysis arose , which is the latest standard in *Clostridia* classification (Moore et al., 2019). *C. difficile* significantly differs from the classical *Clostridia* (Cluster I) with its well-known representative *C. botulinum*. Listed in Cluster XI, it is more closely related but still distinct from a number of bacteria summarized as *Peptoclostridia*, with the representatives *Paeniclostridium sordellii* or *Acetoanaerobium sticklandii* among them (Collins et al., 1994; Monot et al., 2011).

Interestingly, *C. difficile* shows a so-called plastic, or mobile, genome, describing enhanced levels of chromosomal rearrangements. This is leading to an impressive number of several hundred different strains, divided in at least 6 clades (Sebaihia et al., 2006; Knight et al., 2015). The plastic genome manifests itself in a broad range of genome sizes and is possibly driven by adaption to environmental pressure. It is achieved by acquisition of mobile genetic elements such as plasmids, transposons or phages (He et al., 2010; Knetsch et al., 2012; Moore et al., 2019).

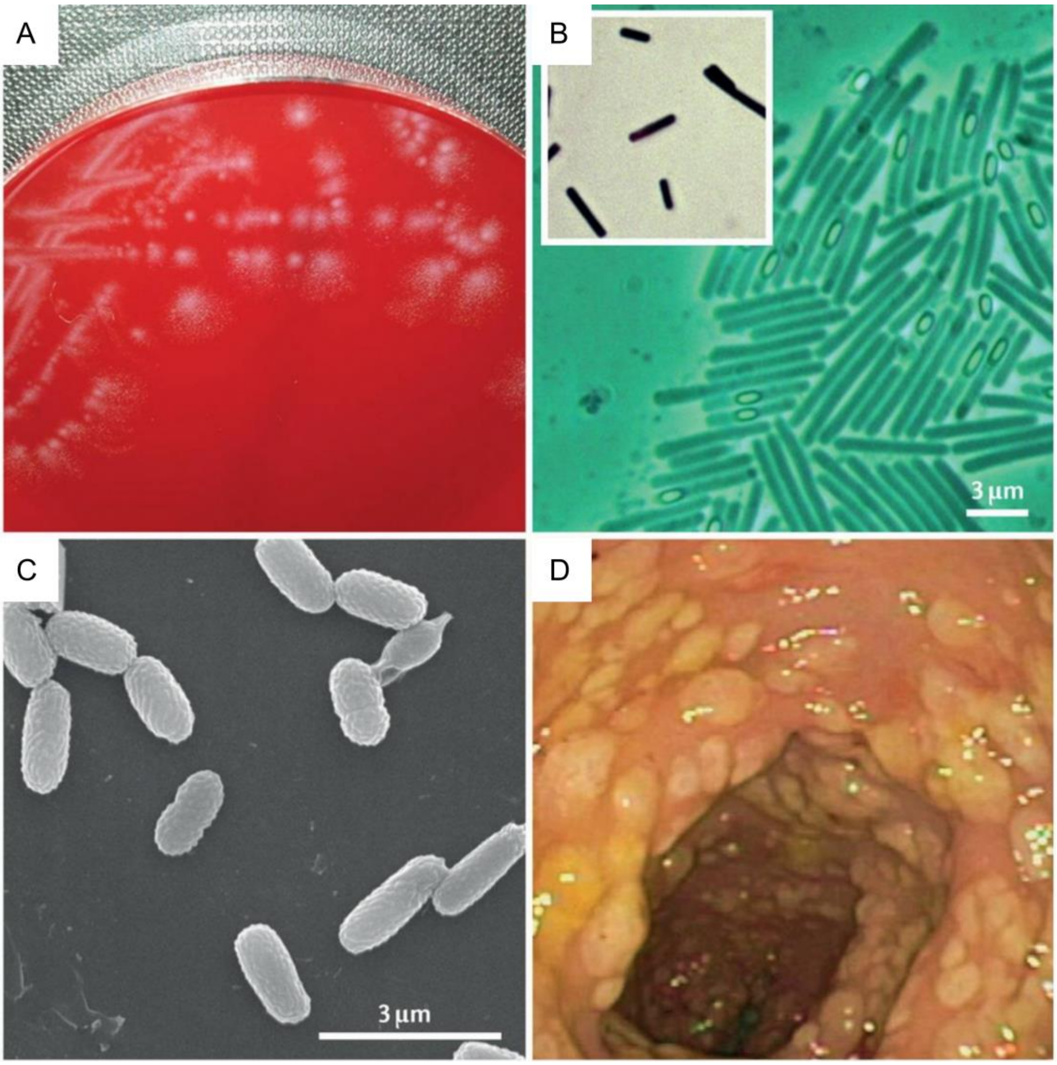


Figure 1.1: Morphology of C. difficile. A) *C. difficile* colonies on a blood agar plate. B) Phase-contrast microscopy image of a *C. difficile* culture. Vegetative cells are visible as elongated rods. Spores are seen as ellipsoids, highlighted by phase contrast. Gram stain shows positive results. C) *C. difficile* spores visualized by scanning electron micrography. D) Endoscopic picture of pseudomembranous colitis caused by *C. difficile*. Healthy colon tissue appears pink, damaged tissue is yellow. Image and caption taken from Smits et al, 2016 (Smits et al., 2016).

C. difficile naturally occurs in human infants (Hall et al., 1935) but is ubiquitously present in our environment. Especially soil and water samples in proximity to animal husbandry show conspicuous contamination and an alarming overlap to virulent strains found in hospitals (Weese, 2010). Particularly, cattle and swine pose a substantial risk, as increasing demand for food enhances concrete danger for regional waves of infections by large-scale farming and environmental contamination (Rodriguez-Palacios et al., 2006, 2007; Yaeger et al., 2007; Hammitt et al., 2008).

Within recent years, *C. difficile* came into focus of research, as 2,824 severe cases of *C. difficile* infections (CDI) were recorded in Germany in 2018, of which 24% were lethal and 31% contributed to death. Severe CDI showed the second highest portion of deceased, only excelled by seasonal influenza (Robert Koch-Institut, 2019). Moreover, an alarming increase of antibiotic resistances was monitored in the recent past, making *C. difficile* to an evident risk for public health issues (Kawata et al., 1984; Barkin et al., 2017; Ong et al., 2017).

1.1.2 Pathogenesis and treatment of *Clostridioides difficile*

Several studies elaborated a clear relationship between antibiotic treatment and a generally increased susceptibility for infectious diseases (Lupp et al., 2007; Kamada et al., 2012; Peterfreund et al., 2012). Unsurprisingly, this relationship also applies to CDI, since it was shown that the risk for CDI increases with application of broad-spectrum antibiotics, such as clindamycin, cephalosporin or ampicillin (Bartlett, 2002; Reeves et al., 2011).

Currently, there is evidence that more than 20% of antibiotic-associated diarrhea (AAD) in humans are verifiably caused by CDI (Bartlett, 1979, 2002; Kelly et al., 1994). Severity of clinical courses of CDI vary, dependent on *C. difficile* strains, patients microbiota, diet and possibly further unrevealed factors (Kelly et al., 1994). The courses are classified as mild, severe or fulminant. Mild courses are characterized by diarrhea and can be treated by oral application of 500 mg metronidazole or 125 mg vancomycin for up to two weeks (Surawicz et al., 2013). Occurrence of fever and strong abdominal pain is observed for severe forms of CDI. Here, vancomycin treatment doses are raised to 500 mg and the application can be rectal (Cohen et al., 2010). In case of poor response, oral dosage of 200 mg fidaxomycin can be considered (Cornely et al., 2014). A fulminant course cannot be precluded in certain cases. It is characterized by severe pain, strong diarrhea and potential development of a toxic megacolon and colon perforation with concrete life-threatening consequences (McDonald et al., 2018). Ultimately, a colectomy can be considered, in which irrecoverable colon sections are removed (Lamontagne et al., 2007).

Interestingly, the occurrence of severe forms of CDI coincidently increased with the identification of a so-called hyper-virulent strain of *C. difficile*, toxinotype III NAP1/027. This strain exhibited an enormously increased level of toxin production compared to other *C. difficile* strains (Warny et al., 2005). Most prominent toxins of *C. difficile* (Tcd) are TcdA and TcdB (Voth et al., 2005; Carter et al., 2012; Awad et al., 2014). Both toxins are encoded in the pathogenicity locus (*PaLoc*) and represent monoglycosyltransferases that glycosylate Rho-GTPases, leading to disintegration of the cytoskeleton, subsequent cell-rounding, the loosening of tight junctions and finally to cell death in the colon (Voth et al., 2005; Steele et al., 2013). Further, it has been observed that TcdA and TcdB also induce levels of proinflammatory cytokines such as TNF-α or IL-8 (Shen, 2012). In addition, the binary toxin, which is encoded by the genes *cdtA* and *cdtB* and is present in some virulent strains, adds another level of virulence (Perelle et al., 1997). Although only a small portion of virulent strains of *C. difficile* carry *cdtA/cdtB*, clinically relevant strains were identified that were TcdA/TcdB-negative and gained their virulence from the binary toxin (Florin et al., 1991). Some hyper-virulent strains, including toxinotype III NAP1/027, show a mutation in the gene *tcdC*, a putative down-regulator of TcdA/TcdB production located in the *PaLoc*, possibly explaining their increased virulence (Mani et al., 2001; Matamouros et al., 2007).

Exact regulation mechanisms for toxin production remain elusive, but researchers have identified a link to the environment of the microbe. In this regard, comparison of low and high-protein diets showed that higher intestinal amino acid content increased the risk of a severe CDI due to higher levels of toxin production. At the same time, a *C. difficile* mutant unable to utilize proline was ineffective to infect mice (Battaglioli et al., 2018).

One important component of *C. difficile*'s central metabolism is the Stickland fermentation, which will be described in the following section.

1.1.3 Stickland fermentation as a central key-player in the metabolism of *Clostridioides difficile*

Stickland fermentation was described in 1934 by Leonhard Huber Stickland as a fermentation of amino acids, coupling oxidative and reductive reactions (Stickland, 1934). It enables proteolytic *clostridia* to use amino acids as sole source of carbon and energy. The reaction can be divided in an oxidative and a reductive arm, which are coupled by a NAD^+/NADH-flux. In more detail, the oxidative reactions are restricted to leucin, isoleucine and alanine as electron donors and comprise oxidative deamination and decarboxylation, both reactions generate NADH. Substrates for the reductive arm are mainly restricted to glycine and, more importantly, proline, which act as electron acceptors. Both reductive deamination of glycine and reductive ring cleavage of proline regenerate NAD^+. Although the product of proline reduction, 5-aminovalerate, is subsequently secreted and is no longer accessible for the metabolism, proline is highly preferred over glycine as electron acceptor and absolutely necessary for growth (Bouillaut et al., 2013).

In general, an upregulation of toxin production is connected to a lack of nutrients and an entry to the stationary growth phase, highlighting the importance for a better understanding of the central metabolism of *C. difficile* (Dupuy et al., 1998). A schematic overview of the proline dependent Stickland fermentation and its role in toxin production is shown in figure 1.2.

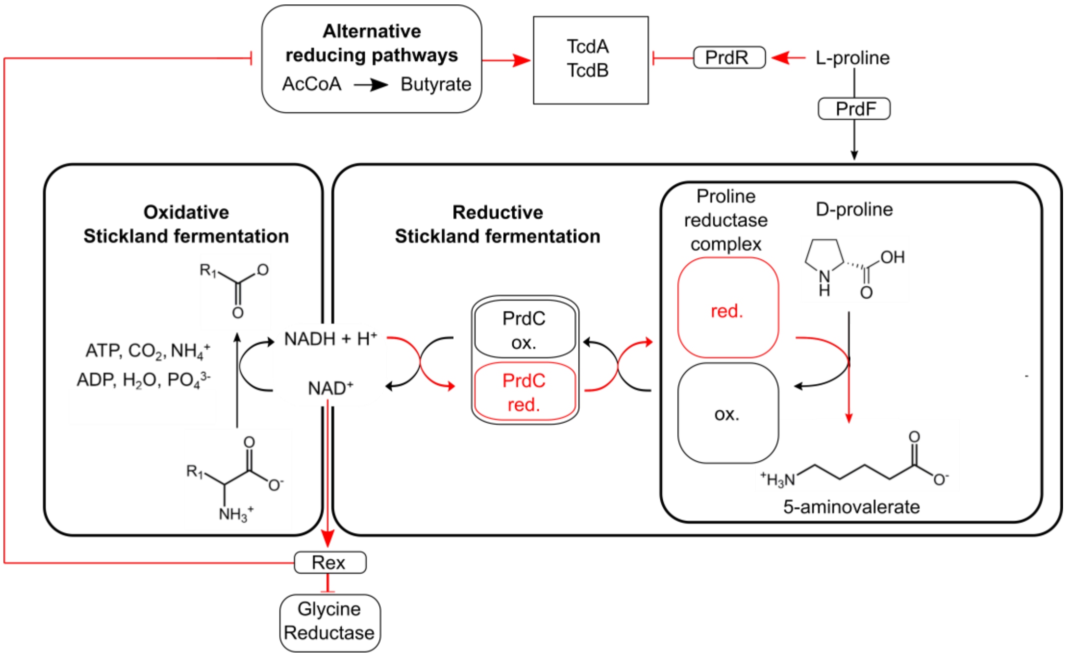


Figure 1.2: Stickland fermentation in its metabolic context. L-proline inhibits TcdA/TcdB production via PrdR. It is transferred to D-proline by PrdF and enters the reductive Stickland fermentation. The proline reductase complex reduces D-proline to 5-aminovalerate under consumption of ATP and NH_4^+. The oxidized proline reductase complex is regenerated by reduced NADH-binding PrdC. Reducing equivalents NAD^+ and NADH couple the reductive arm of the Stickland fermentation to the oxidative arm, where amino acids are oxidatively deaminated and decarboxylated. The NAD^+/NADH-ratio is sensed by the redox-sensing regulator (Rex), which inhibits the glycine reductase and alternative reducing pathways, if proline is present. Alternative reducing pathways would result in accumulation of butyrate, which would again induce TcdA/TcdB production. Conclusively, the presence of proline represses toxin production.

The enzyme fulfilling proline reduction is the proline reductase (PR) (Kabisch et al., 1999). This protein complex was found to be associated with the proton motive force (PMF), possibly explaining the observed preference of its substrate over glycine (Lovitt et al., 1986). Interestingly, there is a notable similarity of the nucleotide-binding PrdC, necessary for regeneration of the PR to its reduced state, to RnfC (Fonknechten et al., 2010 a; Poehlein et al., 2014). RnfC is one subunit of the Rnf complex, essential for generation of a proton gradient and ATP-synthesis, emphasizing the central role of the PR in *C. difficile's* metabolism (Tremblay et al., 2012). *C. difficile* was shown to be thoroughly adapted to the intestinal environment, where typically low concentrations of glucose prevail. It has been shown that *C. difficile* is independent of the consumption of sugars when amino acids are present in the culture media (Ferraris et al., 1990; Jackson et al., 2006). Genomic analysis of closely related species revealed the presence of 12 proteases and 38 peptidases, emphasizing the importance of accessible amino acids during *C. difficile* growth (Fonknechten et al., 2010 b). Recent research discovered and verified a clear link between amino acid availability, the metabolic state of *C. difficile* and toxin production (Martin-Verstraete et al., 2016). In more detail, toxin production is effectively repressed when certain amino acids, such as proline, are available for fermentation (Karisson et al., 1999). When *C. difficile* enters the stationary growth phase, toxin production starts (Dupuy et al., 1998). As long as proline is available, the σ54-factor PrdR stimulates PR expression and is thought to be one responsible factor for toxin repression (Bouillaut et al., 2015). Simultaneously, the redox-sensing system Rex monitors the NAD^+/NADH-ratio in the organism, which is also an indicator for the metabolic state, and represses toxin production (Bouillaut et al., 2019). While the dependencies of metabolic state and nutrient availability on toxin production have been investigated to some extent, the exact relationships and pathways remain unclear and need further, direct detection of proteinogenic interactions.

Yet another context remains enigmatic: surprisingly, it was found that several proline-deficient mutant strains of *C. difficile* were unable to sporulate (Dembek et al., 2015). Indisputably, metabolic fitness and ability for sporulation seem to be closely related. However, no causalities were identified so far.

Therefore, PR must be considered as a central metabolic key-player, essential for growth, with obvious cross-play to TcdA/TcdB regulation and ultimately with a connection to sporulation. Consequently, a profound understanding of PR and its interaction partners will be of immense use in the progress of alternative treatment development.

1.1.4 The proline reductase as a target for structural biology

As mentioned before, the PR is a particularly interesting target of research, as the importance on cellular and metabolic levels has been shown. It reductively cleaves D-proline to 5-aminovalerate (Seto, 1978; Kabisch et al., 1999). However, due to poor accessibility of protein material for research and the small number of homologs, structural information is not available and hence, detailed and valid information about its architecture and mode of action at atomic resolution is missing.

As far as known, the PR complex assembles from the gene products of *prdA* and *prdB*, both located in the PR operon. PrdA matures to PrdAα and PrdAβ by a serino- or cysteinolysis, inducing a backbone rupture at amino acid position 421 (Bednarski et al., 2001). The N-terminal PrdAβ closes with a regular C-terminus, while the C-Terminal PrdAα carries a pyruvoyl moiety at its N-terminus, which is necessary for substrate binding. This maturation does not occur spontaneously but seems to require a chaperone or an appropriate electron donor (Kabisch et al., 1999). The mature PrdAαβ complex forms a complex with PrdB, which is a small protein of 25 kDa and carries a selenocysteine at position 151 (Jackson et al., 2006). PrdAαβ and PrdB form a higher MW complex with yet unknown subunit ratios. Dependent on purification and determination strategies, the described MW of the PR complex varies from 280 kDa (Seto et al., 1976; Jackson et al., 2006) to 870 kDa (Kabisch et al., s1999). Until now, the importance of oligomerization, mechanisms leading to assembly and the involvement of possible interaction partners of PR are unknown. Some results also point towards membrane association of the PR complex (Seto et al., 1976). So far, the PR complex has only been purified from *Clostridia* cell cultures, limiting the amount of pure protein for experimentation.

As the PR complex cannot be structurally predicted by sequence-based methods, experimental structure elucidation is urgently needed to completely understand the mode of oligomerization, interaction and activity. Such findings will enhance the understanding of this metabolic complex and potentially provide new starting points for CDI treatment.

1.2 Topic II: Phospholipases from *Legionella pneumophila*

A number of phospholipases have been found to act as important and highly potent virulence factors of *Legionella pneumophila*. As uncontrolled phospholipase activity is of potential danger for *L. pneumophila* itself, the bacterium has evolved sophisticated mechanisms for export, regulation, inhibition and activation in order to prevent self-inflicted lysis.

1.2.1 *Legionella pneumophila* as pathogen

More than 50 species have been identified in the genus *Legionella*, of which at least 24 are associated to human disease (Newton et al., 2010).

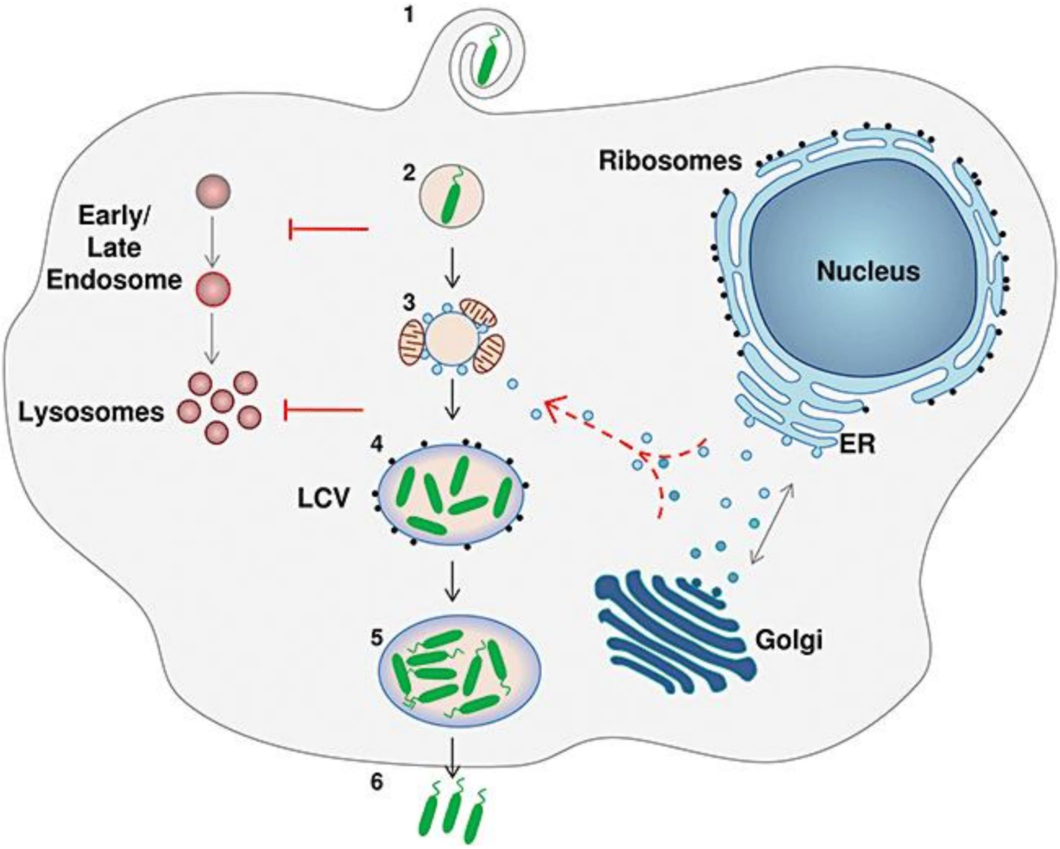


Figure 1.3: Intracellular life cycle of *L. pneumophila*. *L. pneumophila* is taken up by the eukaryotic host cell (amoeba or alveolar macrophage) by phagocytosis (1). Fusion of the *Legionella*-containing vacuole (LCV) with the endocytic pathway (2) and the delivery to lysosomes (3) is prevented by secretion of effector proteins. Recruitment of ER-derived vesicles trafficking to the Golgi fuse with the LCV (3). A replicative vacuole surrounded with ribosomes is formed (4). Inside the LCV, *L. pneumophila* replicates and becomes flagellated (5). Escape from the host cell induces a new start in the infection cycle in neighbouring cells (6). Figure and caption taken from Franco et al., 2009 (Franco et al., 2009)

They naturally occur ubiquitously in fresh water (Fliermans et al., 1981) and replicate intracellularly in *amobae* and *nematodes* (Hales et al., 1999; Swanson et al., 2000). Here *L. pneumophila* modulates its environment and forms a comfortable niche by secretion of effector proteins (Isberg et al., 2009; Hilbi et al., 2011). In a similar manner *L. pneumophila* can also adopt to replication in human lung macrophages upon inhalation of contaminated aerosols (Newton et al., 2010). *L. pneumophila* is the causative organism of Legionnaires' disease, a severe pneumonia, with a number of about 1500 registered cases in Germany in 2018 (Robert Koch-Institut, 2019). The number of infections is continuously increasing since several years, putatively due to changes in the environment. Secreted effector proteins are large in number and some even carry Rho- and Rab-GTPase domains, a unique observation in prokaryotes (Gomez-Valero et al., 2019). Their role in vesicle transport systems shows how *Legionella* manipulate membrane trafficking (Novick et al., 1997). Among secreted effector proteins are also a number of phospholipases, which disturb membrane integrity and produce second-messengers (Schmiel et al., 1999).

1.2.2 Phospholipases as virulence factors

The phospholipases from *L. pneumophila* are of crucial importance for pathogenesis, as they participate in modulation of host cells (Flores-Díaz et al., 2016). Phospholipases can be divided in phospholipases A (PLAs) and lysophospholipases A (LPLA), which hydrolyze the carboxyl ester bonds at sn-1 or sn-2 position in phospholipids and lysophospholipids, respectively (Schmiel et al., 1999). At least 15 genes encoding PLA/LPLAs belonging to three families are found in *L. pneumophila*. Many of these are secreted into the LCV lumen, injected into the host cell or are surface-exposed in order to interact with the host cell (Flores-Díaz et al., 2016; Hilbi set al., 2011; Lang, Hiller, & Flieger, 2017; W. C. Winn & Myerowitz, 1981). Secretion systems are known for many phospholipases and mostly comprise type-2 secretion into the extracellular space for PlaA and PlaC or type-4 injection into the host cell, as exemplified by PatA. For PlaB however, the situation is different. The secretion path used for PlaB is still unclear (Banerji et al., 2008). Most phospholipases are classified as GDSL hydrolases, as they have the GDSL amino acid motif in their active site (Akoh et al., 2004). However, PlaB has a THSTG amino acid motif in its active site and is therefore classified as a THSTG hydrolase (Flieger et al., 2004). A deeper understanding of phospholipase mode of activation and action will hopefully help to define new targets for specific treatment of the Legionnaires disease.

1.2.3 The phospholipase PlaA

PlaA is a 34 kDa phospholipase, which is expressed with a N-terminal secretion signal and has a common α/β-hydrolase fold (ABH-fold) based on sequence-based predictions. It was found to be secreted from *L. pneumophila* by type-2 secretion into the extracellular lumen, were it is activated by ProA-mediated proteolytic cleavage (Flieger et al., 2002; Lang et al., 2017). Interestingly, the enzymatic activity of PlaA changes upon maturation. As a proprotein, PlaA catalyzes a propionylization of cholesterol. Mature PlaA, on the other hand, detoxifies lysophosphocholine (Lang et al., 2012, 2017). It is still a question of current research how PlaA influences the host cell. At present, a role of PlaA in membrane destabilization has been reported (Creasey et al., 2012). A crystal structure of PlaA could potentially increase our understanding of substrate specificity and possibly could enhance the knowledge of the exact biological role of PlaA in *L. pneumophila* pathogenesis. Special attention should be paid to surface-exposed substrate binding sites, the lid domain as an important element for substrate recognition, and closer investigation of the location of the proteolytic cleavage site. Therefore, a crystallographic approach was addressed.

1.2.4 The phospholipase PlaB

The phospholipase PlaB is an essential virulence factor for lung colonization, tissue destruction and intracellular replication of *L. pneumophila*. It localizes to the bacterial surface and shows an extraordinary activation mechanism by deoligomerization. PlaB possesses PLA, LPLA and hemolytic activities essential for intracellular replication in mouse macrophages (Flieger et al., 2004; Bender et al., 2009; Kuhle et al., 2014). PlaB is also crucial for lung colonization and tissue destruction in guinea pig infections (Schunder et al., 2010). It is the only characterized member of a novel PLA family, and homologs are found in several water-associated bacteria including the opportunistic pathogen *Pseudomonas aeruginosa* (Flieger et al., 2004; Bender et al., 2009). Previous work suggested that PlaB is organized in two domains, an N-terminal phospholipase (amino acids 1-~300) connected to a C-terminal domain of unknown function (amino acids ~301-474) that is essential for activity as well (Bender et al., 2009). The catalytic triad Ser_{85}-Asp_{203}-His_{251} of the N-terminal domain is embedded in an uncommon THSTG consensus motive (Bender et al., 2009). The C-terminal domain does not show any similarity to known proteins, but it has been shown that the last 15 amino acids of PlaB are essential for activity although their exact role remains unclear (Bender et al., 2009; Kuhle et al., 2014). PlaB is exported from *L. pneumophila* and associated to the outer membrane of the pathogen, where it is exposed on the surface. However, PlaB lacks export signals as well as lipid anchors or transmembrane helices. Therefore, knowledge of the export mode through inner and outer bacterial membranes is unclear (Flieger et al., 2004; Schunder et al., 2010). PlaB provides the most prominent PLA/LPLA activity in *L. pneumophila*, affecting many different lipids including phosphatidylglycerol and phosphatidylcholine, which is also present in the bacterial membrane. Therefore, enzyme regulation is crucial to prevent cytotoxicity for the pathogen itself (Finnerty et al., 1979; Flieger et al., 2004; Conover et al., 2008). Indeed, PlaB shows a rare activation mechanism that requires protein deoligomerization. It was found that concentrations above 0.1 µM promote inactivation by tetramerization, whereas deoligomerization occurs at nanomolecular concentration and below, where PlaB shows highest activities (Kuhle et al., 2014). However, the detailed mechanism how dissociation of the oligomers mechanistically leads to enzyme activation and the role of the C-terminal domain remains unclear. Detailed molecular understanding could raise opportunities for an early intracellular activation of PlaB, which would induce self-toxicity as a novel treatment option. First crystallographic studies (data not published) showed easy crystal formation, however, the crystals obtained in this earlier work showed poor quality and structure elucidation was impossible. Extensive additive screening and further crystallographic refinement strategies were planned to be applied.

1.3 Principles of integrative structural biology

The term “structural biology” describes all kind of investigations that address the architecture and coupled functionality of biomolecules. Proteins are the most studied examples of research, but also DNA and mixed structures are targeted by structural biology. After the discovery of X-rays in 1895 by Wilhelm Conrad Röntgen, one of many basics for structural biology was set. It took 65 years until the first crystal structure of a protein was published: in 1960, Max Ferdinand Perutz solved the structure of hemoglobin by X-ray crystallography after he solved the phase problem by soaking the crystals with mercury (Perutz et al., 1960). Thenceforth, the newly discovered technique developed rapidly, only in 1990 more than 100 structures were submitted to the Protein Data Bank (PDB), four years later there were 1000 per year already. Until today (spring 2020), more than 140.000 X-ray structures have been deposited to the PDB and thanks to recent approaches such as serial crystallography and fragment screening, the number continues to grow constantly. However, further techniques were developed to determine atomic coordinates in biomolecules. In 1946, magnetic resonance was discovered by Felix Bloch (Bloch, 1946) and the first NMR structure of a protein was determined in 1985 (Williamson et al., 1985). Due to the complexity of the molecules and broad peaks in spectra, the size of the target was limited to approximately 5 kDa (Widmer et al., 1989). The peptide or protein size still matters in NMR spectroscopy, although resolution was pushed by stronger magnets and hence improved significantly. In order to detect carbon and nitrogen in NMR spectroscopy, the atoms need to be substituted by isotopes that have a spin unequal to zero, which makes sample production costly. The application is still limited to proteins smaller than 30 kDa today. However, advantages of the NMR technique are the direct detection of interactions of other atoms or molecules with the labelled target and needs of crystallization are omitted. In respect of molecule size, X-ray crystallography has its upper limits between 150 and 200 kDa per molecule. Structures of (homo)oligomers of higher total MW were successfully determined, but their number is comparably low, as well. Here, another technique stands in that is capable to capture supercomplexes. Electron microscopy (EM) was developed in 1931 (Ruska, 1987) and first images of vitrified samples at cryo-conditions were collected in 1987 (Dubochet et al., 1987). Since its development, limitations in EM are found in very small specimens. An important breakthrough was achieved by the determination of the hepatitis B protein core complex with 5 MDa to a resolution of 7.4 Å (Böttcher et al., 1997). Until recently, the resolution of this complex was pushed to 2.66 Å (Böttcher et al., 2018). Currently, the highest resolution was achieved for the hexameric glutamate dehydrogenase with 1.8 Å (Merk et al., 2016) and also molecule size decreases continuously, filling existent gaps between crystallography and cryo-EM. Due to chronical advance and accessible infrastructure, X-ray crystallography is still the dominant technique in structural biology, contributing to almost 90% to the PDB, as shown in figure 1.4 (Berman et al., 2000).

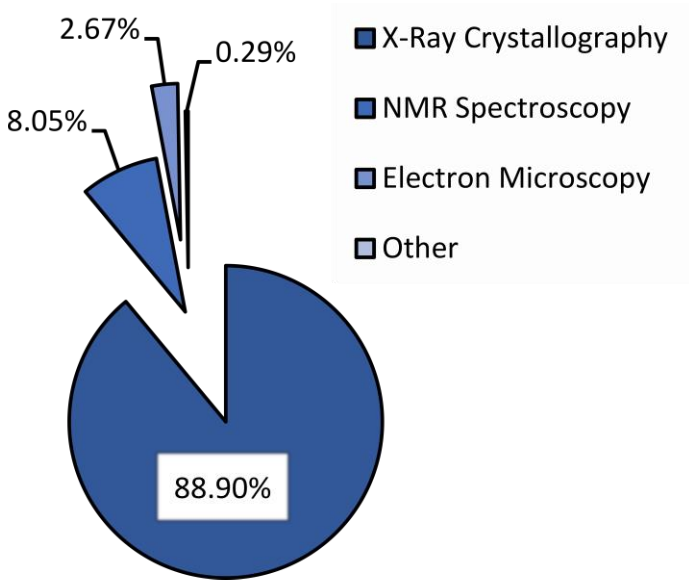


Figure 1.4: Macromolecular structure data of the Protein Data Bank (PDB). The diagram shows the portions of structures solved by different experimental methods (X-ray crystallography, NMR spectroscopy, electron microscopy and other methods). X-ray crystallography is the most prominent technique, represented by almost 90% of all deposited structures. NMR spectroscopy follows with a larger subset of 8%. Since recently, the number of EM structures increased tremendously, currently providing 2.7% of the deposited structures.

Overcoming size limitations, further aspects hinder certain biological macromolecules from being structurally described. Membrane proteins are in general hard to access, as the biological membrane needs to be reconstituted for sample stability. However, development of nanodiscs enables EM to visualize membrane proteins as single particles (Hauer et al. 2015; Denisov and Sligar 2016; Sun and Gennis 2019).

As a complementary method, SAXS still seems to be underestimated. The sample requirements are comparably low and despite the existing ambiguity in the well-known visualization of the results, SAXS can provide precious and precise information on molecular weight, particle size distribution and hence is suited to prove biomolecular complex formation.

The term "integrative structural biology" describes the combination of all techniques mentioned above. Integrative structural biology generates synergistic effects by circumventing weaknesses and exploits strengths of NMR, X-ray crystallography, EM and SAXS. In the following section, the different methods are briefly introduced.

1.3.1 X-ray crystallography

Protein crystals show properties that are highly advantageous for high resolution data collection. In a crystal, the molecules pack in a large number into a three-dimensional arrangement, established by repetitive molecular contacts on the protein surface as indicated in figure 1.5. Therefore, compared to a single molecule, the signal to noise ratio for scattered light is much higher.

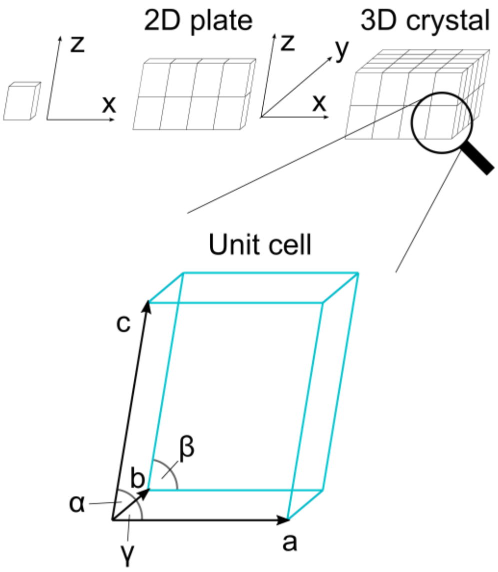


Figure 1.5: The crystallographic unit cells. A crystal packs an arbitrary number of unit cells in a repetitive manner to form two-dimensional crystal plates and three-dimensional crystals. The cell is defined by the cell axis a/b/c and angles α/β/γ.

The diffraction itself results from destructive and constructive interferences, caused by the interaction of the X-ray beam with a repetitive pattern of molecules in crystals and is detectable to high diffraction angles, consequently leading to potential atomic resolution (Rupp, 2009). At the same time, crystallization itself causes one of the major restraints of this technique. A protein that gains its biological function from its flexibility is fixed in a solid state. Not every conformation is equally likely to crystallize and some conformations might not be depicted at all. Hence, a crystal structure is always a snapshot of a certain protein conformation. However, structural information is of high value in protein sciences and subsequent soaking experiments, structure guided mutagenesis and other experiments enable scientist to decipher molecular reaction mechanisms (Buck, 2003).

Crystal growth is most commonly achieved in a vapor diffusion experiment. Here, small volumes of a protein solution are mixed with equal volumes of a precipitation solution. Common precipitants are high concentrations of salt, alcohol or polymers such as polyethylene glycol (PEG) that decrease protein solubility. In many cases, the pH of the precipitation formula is kept constant by the addition of an organic buffer compound. Certain reagents that can be essential in small amounts for crystallization are referred to as additives. The mixed droplet is placed in a sitting well, next to a reservoir well that holds the precipitation solution in a large excess. The whole entity is sealed and vapor diffusion occurs, meaning that water evaporates from the droplet and condenses in the reservoir, due to dilution of the precipitant in the crystallization droplet. This behavior increases protein and precipitant concentration slowly but gradually. In case of matching conditions, eventually protein crystals may grow from the solution as solubility limit of the protein is exceeded (Rupp, 2009).

$$S = s - s_0 \tag{1}$$

$$|S| = \frac{1}{d} = \frac{2sin\theta}{\lambda} \tag{2}$$

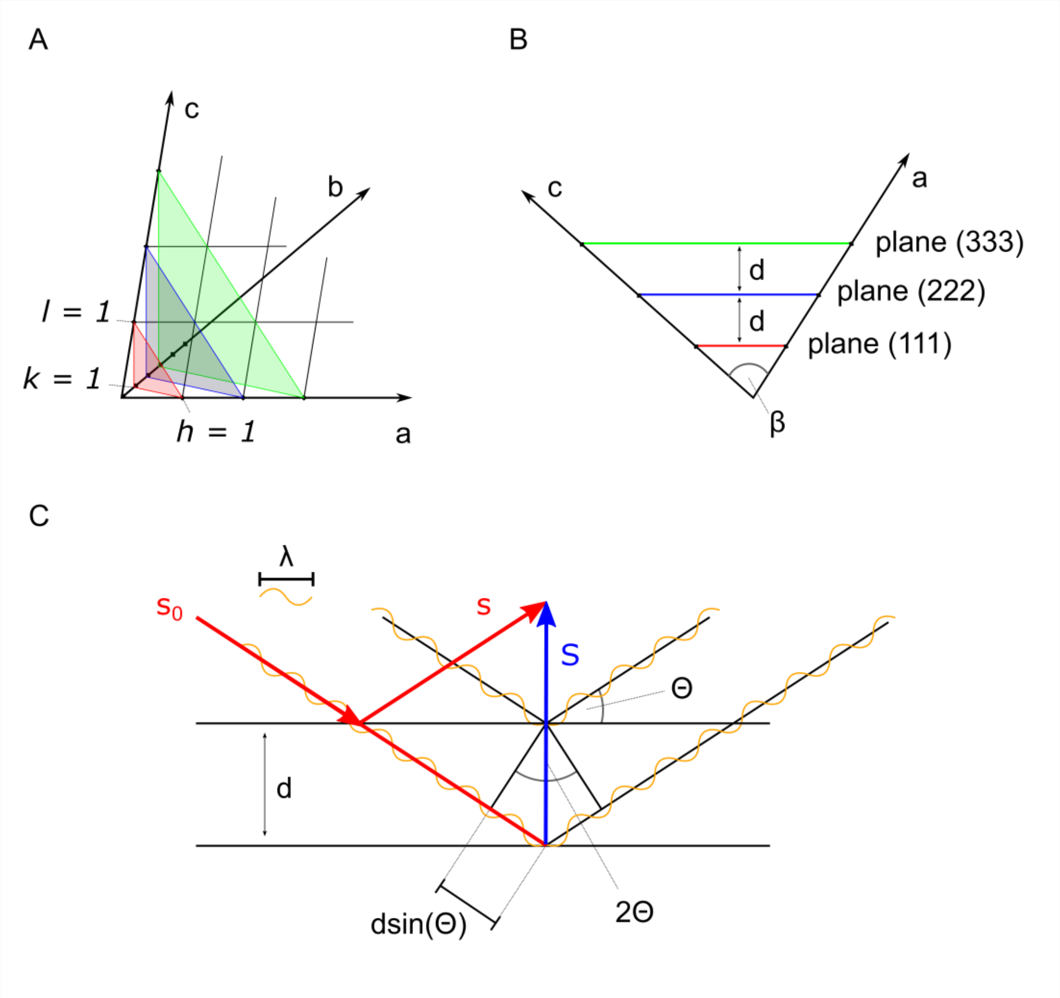


Figure 1.6: Bragg's Law. Diffraction of X-rays by a crystal. A) The crystallographic unit cell generates a reciprocal lattice which generates lattice planes (red, blue and green triangles). B) Lattice planes are exemplified by (hkl) equals (111), (222) and (333). The distance d between lattice plains correlates with achievable resolution from crystal diffraction. C) Two identical X-rays (orange wave) diffracted by reciprocal lattice planes (black). The travel distance difference is defined by 2dsin(Θ) and generates a phase shift between the X-rays. When 2dsin(Θ) is an integer of λ, Bragg's Law is fulfilled and constructive interference of the waves occurs, resulting in a detectable diffraction spot. The scattering vector S (blue) is defined by the vector difference of s and s_0 (red).

As finding the correct crystallization condition is a unique procedure for every protein and may be elaborate, protein crystallization is usually performed in a 3 x 96-well format and with machine support.

Crystals are flash frozen in liquid nitrogen (LN_2) after addition of cryo-protectants that promote vitrification. Data collection is performed by X-ray exposure. The crystal is rotated in the beam to project all reflections possible on a detector. Diffraction occurs when Bragg's Law is fulfilled, as shown in figure 1.6 (Bragg, 1913).

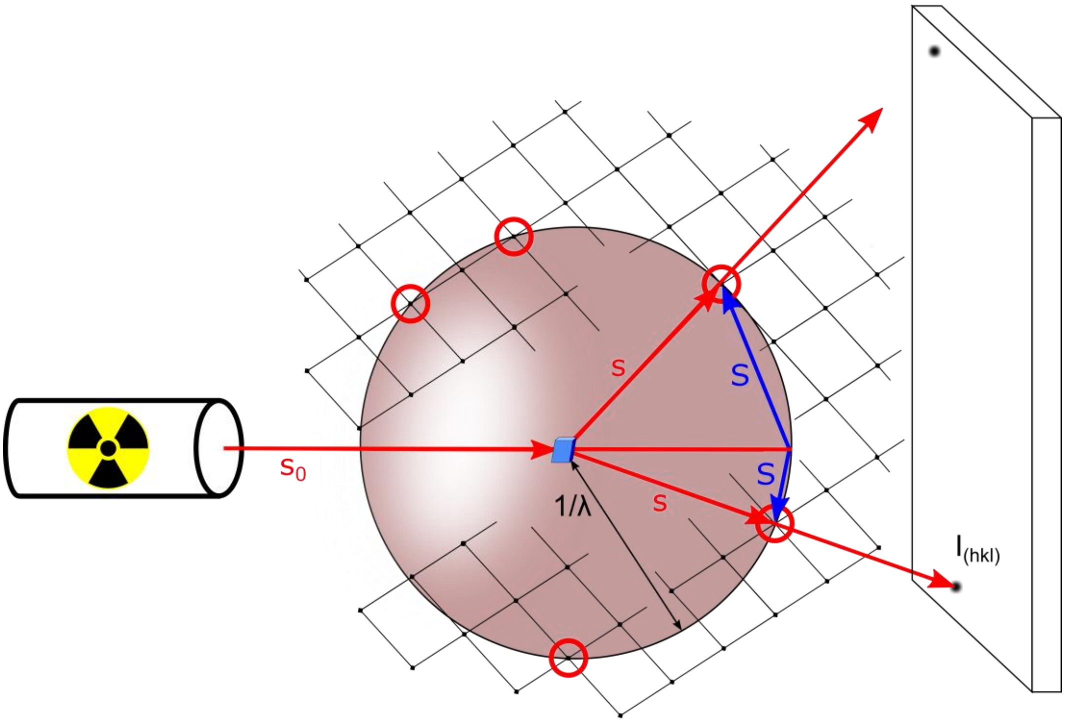


Figure 1.7: Ewald's sphere. The illustration shows how the Ewald's sphere is defined by a radius of 1/λ. With its origin in (000) it illustrates that diffraction occurs when reciprocal lattice points meet the shell of the sphere. Such events are highlighted with red circles. Scattering of the incident beam s_0 results in scattered vectors s (red). The vector difference S (blue) is unique for each reflection, captured on a detector. The detector collects position and intensity of each individual spot by rotation of the crystal in the X-ray beam.

Rotating the crystal in the beam during data collection, causes temporary and angle dependent reflection detection. The occurrence of reflections can be understood by the implementation of the Ewald's sphere with a radius of 1/λ and an origin in the crystal center. When a lattice point of the reciprocal lattice touches the Ewald's sphere, a reflection is observed (see figure 1.7).

The collected data is then forwarded to data processing. In a first step, the data is indexed. Here, the diffraction angles and relative positions of reflections to each other are determined. From mirror symmetries in the diffraction pattern, the rotational symmetry in the crystal and hence the space group can be determined. The distance of diffraction spots in h/k/l-directions corresponds to the inverse of the unit cell lengths a/b/c. Therefore, reflections that lie close together in *h*-direction correspond to a long unit cell axis in a-direction. This way, the unit cell parameters a/b/c and α/β/γ are determined. Systematic absences of reflections, where the structure factor is zero, indicate the presence of a screw axis, a combination of rotation and translation, projecting two molecules on each other.

The subsequent integration step determines spot intensities $I_{(hkl)}$, necessary for structure determination. $I_{(hkl)}$ is proportional to $|F_{(hkl)}|^2$. However, the phase information $i\varphi_{(hkl)}$ cannot be recorded. This leads to the so-called crystallographic phase problem, which describes that the electron density in real space $\rho_{(xyz)}$ cannot be determined by the structure factor equation (x), as the phase and the density are two variables that are missing.

$$\rho_{(xyz)} = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{(hkl)}|\, exp\left[-2\pi i(hx+ky+lz)+i\varphi_{(hkl)}\right] \tag{3}$$

Transformation of waves with phase and amplitude information are converted to the complex number system. Here, the wave amplitude is represented by the absolute value and the phase is given as the angle of a vector. This procedure makes wave addition easier and intuitively possible by vector addition. The structure factor $F_{(hkl)}$ is a complex number and can therefore be simplified by illustration in an Argand diagram, as shown in figure 1.8 (Rupp, 2009).

$$F_{(hkl)} = A + iB \tag{4}$$

where the measured value |F| is

$$|F| = \sqrt{A^2 + B^2} \tag{5}$$

and

$$A = |F| cos\Phi \tag{6}$$

and

$$B = |F| sin\Phi \tag{7}$$

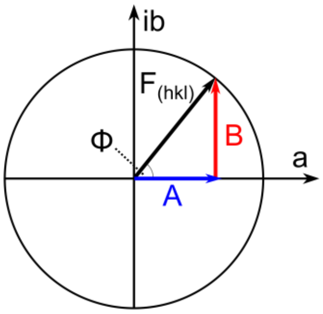


Figure 1.8: Argand-diagram. Complex numbers can be illustrated as a vector with an amplitude and a rotation angle. In X-ray crystallography, the amplitude correlates with measured spot intensity. The rotation angle represents the phase.

The aim is to determine $\Phi_{(hkl)}$, which can be achieved by molecular replacement (MR) or experimental phasing. For MR a homologous structure must be available. By rotational and translational functions, model molecules will be fit in the asymmetric unit (ASU) in an estimated number. When experimental and theoretical structure factors converge, phases from the model can be used (Rupp, 2009).

If MR fails, due to a missing model or other crystallographic issues, experimental phasing offers another option to determine crystallographic phases. Here, anomalous scatterers, such as heavy metals, are incorporated into the crystal and a dataset is recorded exploiting scattering differences for phase determination as shown in figure 1.9 (Rupp, 2009).

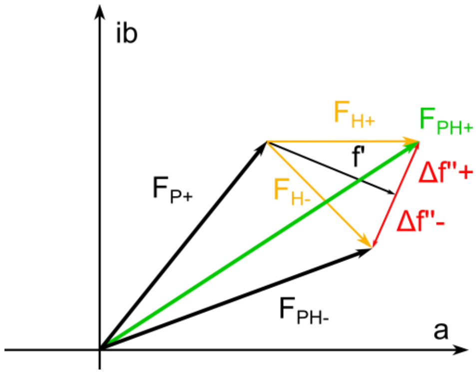


Figure 1.9: Structure factors. In native data collection the structure factor amplitudes for protein (P) are symmetrical. Therefore, the values for F_{P+} and F_{P-} are equal (Friedel pairs). Upon incorporation of heavy metal scatterers (H), the symmetrical, native structure factor amplitudes for F_{P+} and F_{P-} are modified by the anomalous atomic scattering factors f' and Δf''. This results in structure factor amplitudes F_{PH+} and F_{PH-}, which are unequal in value. The structure factor difference can be used to determine phases experimentally.

Once initial phases for a dataset are obtained, model refinement aims at lowering the difference for observed structure factors to model-based calculated structure factors. Evaluation includes electron density interpretation, peptide bond angles for structural elements or rotamer conformations, among others.

1.3.2 NMR spectroscopy

NMR spectroscopy is another technique of structural biology with its own strengths and weaknesses. In a static magnetic field, NMR sensitive nuclei find preferred z-orientation. This behavior is called the Zeeman effect. When a second magnetic field is applied with an oscillation equal to the nuclei's Larmor frequency and perpendicularly to the static field, a reorientation to the x,y-plane is enforced. This magnetic field is applied in pulses of varying lengths. When the pulse is switched off, the nuclei fall back to their initial state and the emitted signal, which depends on the chemical environment of each single atom, can be recorded by the free induction decay (FID). To be NMR-sensitive, the quantum spin number of the atom may not equal zero, which is dependent on the number of protons and neutrons in the atomic core. The most relevant examples for the spin number in proteins are given in table 1.1 (Cavanagh et al., 2007).

Table 1.1: Selection of elements and their accessibility for NMR spectroscopy.

Isotope	Spin I	Natural abundance (%)
^{1}H	½	99.98
^{2}H	1	0.016
^{12}C	0	98.9
^{13}C	½	1.108
^{14}N	1	99.63
^{15}N	½	0.37

The Nuclear Overhauser Effect (NOE) arises from inter-atomic interactions and is detectable for atoms with an average distance of up to 5 - 6 Å (Markwick et al., 2008). The chemical shift for an atom depends on its environment, which is why, for example, a differentiation between backbone protons of helices and sheets becomes possible (Case, 1998; Wishart et al., 2002). Unlike crystallography, each chemical shift signal contains information on only one nucleus, making this technique meaningful for direct interaction studies. Visual comparison of apo- and holo-spectra enables identification of residues that show a chemical shift upon interaction with another molecule or a change in the environment. NMR spectroscopy is usually performed in solution, minimizing preparation time and costs.

A 1D ^{1}H NMR spectroscopy, performed as shown in figure 1.10, depicts a quick validation of sample quality, giving information on protein size, oligomerization and protein stability.

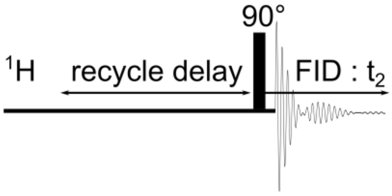


Figure 1.10: Pulse sequence for 1D proton NMR spectroscopy measurement. After a recycle delay, a 90° pulse is applied. The detector senses maximal signal during the FID.

Due to the large number of coupling protons in a protein, a simple 1D proton spectrum results in an extremely complex spectrum where single signals cannot be assigned to their originating protons. However, this experiment is an essential component of NMR studies and is typically performed in an early stage of a project, as the peak width already provides precious information on the sample quality and the protein oligomerization state.

A more comprehensive method is multidimensional NMR spectroscopy. Typical 2D experiments show a more complex pulse-sequence, shown in figure 1.11. Usually, they consist of a first preparation pulse, followed by an evolution time t_1 and a second mixing pulse, followed by data acquisition time t_2. Correlated spectroscopy (COSY), nuclear

Overhauser enhancement spectroscopy (NOESY) and total correlation spectroscopy (TOCSY) are typical 2D experiments that do not require heteronuclear labelling. HSQC, the so-called protein fingerprint, requires ^{15}N-labelling.

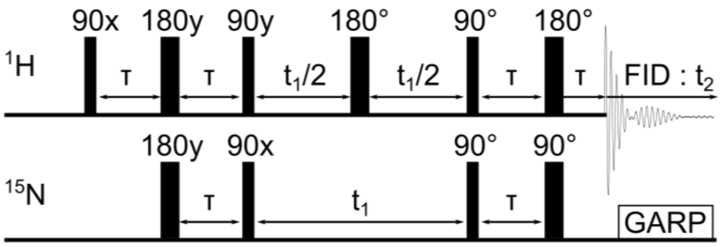


Figure 1.11: Pulse sequence for 2D HSQC NMR spectroscopy measurement. 1H and 15N experiments are performed iteratively.

The globally optimized alternating phase rectangular pulse (GARP) is used for efficient decoupling of certain nuclei. Therewith, spectra complexity is drastically reduced and the results become interpretable. Successful data collection allows signal assignment to individual residues, a non-trivial task. However, automated assignment and structure determination was developed during recent years, potentially enhancing the availability of NMR spectroscopy to a broader range of researchers (Meiler et al., 2003).

Besides costly and time-consuming ^{15}N or ^{13}C labelling procedures, one main difficulty in NMR-based structural biology is the size limitations to smaller molecules. The relaxation, base for signal detection, occurs with enhanced speed by increasing size of the molecule. This shortens the acquisition time and leads to broader and more overlapping signals. However, increasing magnetic strength widens the capable spectrum of protein size and until recently, useful data on proteins as big as 70 kDa could successfully be recorded (Xu et al., 2006; Purslow et al., 2020).

Data acquisition is usually performed between 20°C and 30°C in a liquid environment. This enables NMR spectroscopy to study molecular dynamics, folding or unfolding of a sample and the direct interaction of target proteins to ligands. Further, circumventing poor crystallization behavior of a high number of disordered or flexible proteins fully justifies costly and laborious sample preparation.

1.3.3 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) provides a relatively easily accessible method to determine a trustworthy value for the MW of a protein in both, batch mode and online mode when coupled to a SEC system.

SAXS is the measurement of scattering differences between the target solubilized in buffer and the buffer itself. Scattering angles and intensities are low, as the interaction of light and matter is not constructively enhanced like in crystallography. Thus, the background scattering should be kept at a low level to yield a reliable scattering profile for the target after data subtraction. High concentrations of additives such as PEG or glycerol should be avoided, just as high concentrations of salt. The background buffer must match the sample buffer down to the smallest detail, which is why sample dialysis must be performed after a protein concentration step for a batch measurement. In SEC-SAXS, a section from the scattering profile before elution of any material is adducted for background subtraction. The experimental setup is shown in figure 1.12. An X-ray beam passes a sample in a flow cell. A light scattering detector is placed behind the flow cell. Due to small scattering angles, the distance from the cell to the detector is much larger, compared to a crystallographic diffraction experiment. As the molecules are randomly oriented in solution, the scattering intensity over the scattering angle is recorded. Distinct diffraction spots cannot be observed, as a crystallographic lattice and therefore wave interference is missing.

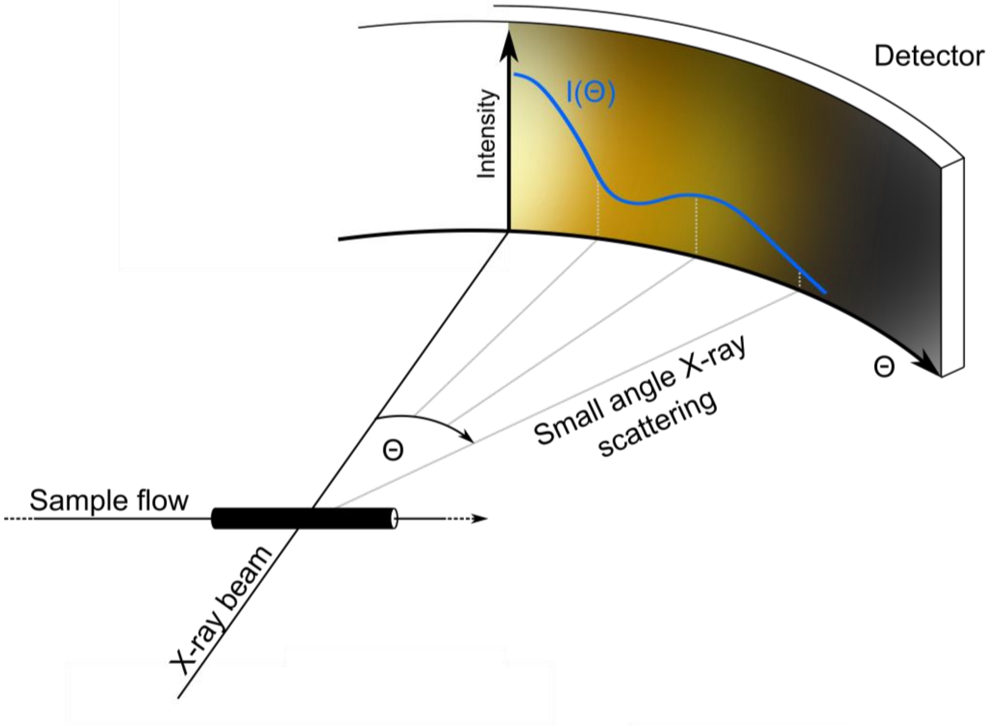


Figure 1.12: Experimental setup SAXS experiments. The incident X-ray beam is scattered by passing the sample. Angle-dependent scattering intensities are detected. The scattering curve is further processed.

The averaged and subtracted scattering curve provides information of particle size and shape. The size of a scattering particle is determined from the Guinier analysis. In analogy to MALS, described in chapter 3.6.5, the radius of gyration Rg is determined from the intensity of scattered light at low scattering angles. In SAXS, the number of data points is much larger than in MALS, due to usage of a detector plate, for example the PILATUS3 X 2M (Dectris, Baden-Daettwil, Switzerland), a hybrid photon counting detector with almost 2.5 megapixels, which collects intensities over a wide range of small X-ray scattering angles. When presenting the data as a Guinier plot with ln(I_s) over s^2, where s is the scattering vector, dependent on the scattering angle 2Ɵ and the wavelength λ,

$$s = \frac{4\pi sin\theta}{\lambda} \tag{8}$$

a linear range in the low angle scattering region is suitable for interpolation of the intensity of the scattered beam at an angle equal zero, I_0.

Following the Guinier approximation, the Rg can be derived from I_0 and be used for MW estimation.

$$I(0) = \frac{I(s)}{e^{\left(\frac{-s^2Rg^2}{3}\right)}} \tag{9}$$

The distance distribution describes the probability p of finding a point, for example. an atom, with a distance r to any other point in the system, dependent on the intensity I and the scattering vector s.

$$p(r) \ = \ r^2 \int_0^\infty I\,(s)\frac{sin(sr)}{sr} 4\pi s^2 ds \tag{10}$$

Plotting p(r) over r provides the pair distance distribution function (PDDF) allows estimations on the particles size and shape. Examples for theoretical bodies are shown in figure 1.13.

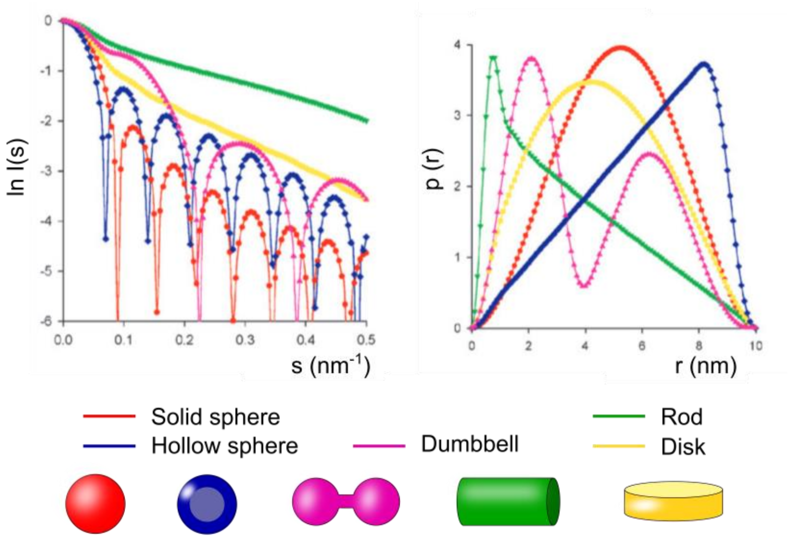


Figure 1.13: SAXS profile and pair distance distribution function. Both diagram depend on the scatterers size and shape (adopted from Svergun & Koch, 2003).

The r-value at the point at which the PDDF converges to zero is used to determine the maximal diameter D_{max} of the scatterer. D_{max} depicts an important value for validation of complementary experiments and is also used for MW estimations. Further, the PDDF serves for the generation of a 3D bead model, where the beads represent hypothetical single scattering points. The model results in a low-resolution envelope, which is generated to fit the experimental data as closely as possible. This envelop can give precious information on shape, oligomerization state and complex assembly of proteins which are inaccessible for crystallization or other structural biology techniques, especially when partial substructures are available (Pauw, 2013).

Untrustworthy data can quickly lead to misleading interpretations when collected on samples of poor quality. Therefore, it is of crucial importance to prepare samples carefully and to pay attention to measurement requirements. Inhomogeneous particle size distribution due to agglomeration, degradation or a dynamic oligomerization equilibrium will result in signal overlap and an averaged scattering profile, making a useful data interpretation impossible. Hence, a high homogeneity and protein stability is required for batch measurements. Targets that show agglomeration upon thawing, a dynamic oligomerization equilibrium or other peculiarities, should be considered to be tested in a more time consuming SEC-SAXS mode (Pauw, 2013).

1.3.4 Electron microscopy

Hans Busch was the first to use electromagnetic lenses to influence a beam of electrons. With this knowledge, electron microscopy (EM) was developed more than 80 years ago by Ernst Ruska (Ruska, 1987). Since then, the technique developed rapidly and generated a set of variations, like transmission EM, electron tomography or electron scattering. Within the last years this field was opened to structural biologists, offering new perspectives on proteins that could not be studied in their structure before. EM is the only method in structural biology that delivers direct images of the sample.

A suitable source produces electrons that are accelerated in an electromagnetic field and used to display an image of a specimen. As electrons cannot be redirected by classical lenses like optical light, solenoids are used that generate tunable electromagnetic fields, suitable to take influence on the electrons speed and course. When the electrons hit the specimen, they are scattered elastically and inelastically. Solenoids will refocus the scattered electrons that will be detected by a CCD or CMOS camera, whose development mainly contributed to the generation of sharper images.

Confining on proteinogenic samples, transmission EM (TEM) is widely used, where the electrons pass through the specimen. Conventional negative stain TEM is used to achieve low resolution information, typically limited to around 20 Å for decent sample quality. Here, the proteins are blotted on a grid and embedded in a layer of heavy metal salts, such as uranyl acetate, ammonium molybdate or phosphotungstate to enhance the contrast between background and protein. The specimen must be dried, as residual water would evaporate in the vacuum and destroy the specimen. The main drawbacks of negative stain TEM are rooted in the sample preparation itself. Heavy metal ions are prone to react with proteinogenic side chains such as cysteines and may cause denaturation. Further, water withdrawal and the prevailing low pH that is necessary for the most common negative stain reagents must be tolerated by the sample. Hence, sample stability depicts the major issue in negative stain TEM. On the other hand, the high contrast enables data acquisition at smaller and cheaper instruments compared to cryo-EM and sample preparation and storage is kept at an economical level. Occasionally, negative stain TEM data serves as a proof of principle and a pre-stage of cryo-EM.

The idea of embedding a target in a layer of vitrified ice is as old as EM itself and is referred to as cryo-EM (Dubochet et al., 1985, 1987). Referring to proteinogenic targets, the clear advantage in comparison to negative stain TEM lies in the presence of vitrified water. This way the protein can be kept in a native-like environment with optimized buffer conditions and the hydration shell is preserved, stabilizing the sample. While advantageous for protein stability, the very same procedure causes extremely low contrast, as the electron density for proteins and surrounding buffer differ only slightly. Hence, cryo-EM can only be performed with large microscopes with 200 or 300 keV acceleration voltage and sensitive cameras.

Within the last years, the MW limit of addressable proteins was lowered continuously from the megadalton size range down to a few hundred kilodalton. Simultaneously, achieved resolutions improved quickly. Mostly, these improvements are based on the development of direct single electron detectors, providing smaller pixel size and higher sensitivity. Also, processing algorithms improved, exploiting eventual symmetries or differentiating between several conformational states. The vast development in cryo-EM map quality is referred to as the “resolution revolution” (Kühlbrandt, 2014).

However, trustworthy resolution determination in EM remains difficult. Most commonly the Fourier Space Correlation (FSC) is used, comparing two individually refined maps from the same dataset. Applied thresholds for resolution determination have encountered various changes. While in early studies, an FCS value of 0.143 was considered to determine the resolution, its value had been raised to 0.5 in more recent studies (Afonine et al., 2018).

The biggest strength of cryo-EM lies in the possibility to encounter large protein complexes which are highly unlikely to crystallize. Intrinsic flexibility does not hinder reconstructions *per se*, however final resolution may suffer. In the nearer future, cryo-EM could potentially be used to circumvent difficult crystallization of medium-sized proteins, classified as non-crystallizable. The development of nanodiscs even enabled the visualization of membrane-incorporated proteins (Hauer et al., 2015). However, specimen preparation can be time consuming and costly and even exceed laboratory efforts required for crystallization.

The most prominent hurdle in cryo-EM is described by a poor reproducibility of blotting procedures, despite sophisticated automatization techniques. The ice thickness has a strong influence on sample quality. Thicker ice causes higher background and lower contrast, impeding successful data acquisition. On the other hand, a layer too thin leads to protein denaturation at the air-water-interface. At the same time, particles should not overlap and be distributed equally in randomized orientations (Drulyte et al., 2018). A common cause of reduced data quality is a preferred sample orientation, presumably due to interactions at the air-water-interface. Preferred particle orientations can be addressed by tilting the specimen or supplementation of detergents (Zi Tan et al., 2017).

Initial particle picking is performed manually and hence may introduce a bias. It clearly requires scientific experience. Subsequent automated particle picking results typically in several hundred thousand or million single particles, which are averaged and used for a 3D reconstruction. However, a large number of particles may be excluded from calculations, if not fitting to averaged classes. Still, all particles present in the sample potentially present relevant biological conformations. Experimental approaches are used frequently to favor certain conformations in order to reduce computing efforts. Typically, small molecules known to interact with the target or their derivatives are added to capture different states of the target (Scheres, 2016). Beside additive screening, flexibility and stability of the target can further be decreased by introducing crosslinks prior to or during an additional separation step (Stark, 2010; Adamus et al., 2019). However, data interpretation of crosslinked samples must be taken with caution, as the created complex may be artificial and the biological relevance may not be given.

The immense efforts, on the other side, are countered by implementation of automated processes. Those are realized in the laboratory routines by automated blotting procedures or support data analysis, for example by *e pluribus unum* (EPU) workflows. In general, data acquisition covers a range of defoci for higher contrast and images that are corrected by contrast transfer function (Ctf).

2 Research Objectives

In the recent past, a spirit of optimism has emerged in almost all fields of biology. Many developments in the field of cryo-EM have opened new perspectives for the study of large protein complexes and cellular machineries, recalling their long-known but nevertheless neglected importance in biological studies. The study of new emerging infectious diseases demands an integrative approach in structural biology, which includes techniques such as NMR spectroscopy, X-ray crystallography, SAXS and cryo-EM combined with biochemical characterization. This has helped to generate synergy in experimental outcomes for the most pertinent questions in infection biology, some of which being: How do pathogens infect humans? How to mitigate different courses of disease? Can science offer contemporary agents against emerging antibiotic resistances?

Within a rapidly changing world, novel threats for human health emerge. Especially in developed countries, where antibiotic doses are prescribed recklessly and personal responsibility of patients is low, antibiotic resistances are a common issue. Of particular importance are nosocomial infections caused by opportunistic pathogens turning into a dangerous complication for patients with a compromised immune system, as it occurs after surgeries or application of broad-spectrum antibiotics. One important example is the opportunistic pathogen *Clostridioides difficile*. This microbe has the exceptional ability to adapt and annihilate microbial gut flora in humans. Its ability to overgrow and dominate the intestinal microbiota is mainly linked to the *clostridia*-specific Stickland fermentation and the main cause of antibiotic associated diarrhea and severe colon inflammation. This pathway uses amino acids as a source of carbon and energy and is an important regulatory mechanism for sporulation and toxin production. It is therefore of immense importance for *C. difficile* and its toxicity, but has been insufficiently investigated until now. An indispensable component of this pathway is the proline reductase complex, which forms a high molecular weight complex assembled from smaller subunits (PrdAα, PrdAβ and PrdB). This attribute makes it an attractive target for an integrative structural biology approach. A deeper understanding of the structural composition of the proline reductase complex can potentially provide innovative targets for treatment *of C. difficile* infection. Thus, the subunits PrdAα, PrdAβ and PrdB of *C. difficile* proline reductase were investigated using a broad spectrum of structural biology techniques, such as negative stain TEM and cryo-EM.

Apart from hospital-acquired infections, new modes of airborne infections occurred within the last decades that are spread by man-made facilities. A pathogen causing such infections is pneumonia-causing *Legionella pneumophila*, which gained notoriety since its first recorded outbreak in 1976. Since then, yearly numbers have increased slowly but steadily. Climate change and increasing use of air conditioners seem to be the main possible reasons for yearly spikes in such infections. *L. pneumophila* adapts to a replicative state inside human lung macrophages by modulating their immediate microenvironment. Among other factors, this is enabled by a number of phospholipases, depicting highly potent virulence factors that affect host cell membrane integrity and release secondary messengers.

One example of such a phospholipase expressed and secreted in *L. pneumophila* during its infection cycle is PlaA. While its exact role in the pathogen's lifecycle remains unclear, it has been shown that PlaA must undergo proteolytic maturation, carried out by another, also secreted, effector protein, ProA. Thus, it is an important event and protein X-ray crystallography was used in this study to investigate the influence of the maturation of PlaA on the protein structure and the role of a disulfide bond in its enzymatic activity.

Another phospholipase, exhibiting the most prominent PLA activity, is PlaB, which shows interesting properties in terms of its localization, oligomerization and regulation. The origin for PlaB's unusual properties is thought to be encoded in its uncharacterized C-terminal domain. In this thesis, the structural characterization of PlaB was addressed by X-ray crystallography, paired with biochemical characterization and structure-guided mutagenesis with the aim to increase understanding of its unique properties in regulation and its role in *L. pneumophila* pathogenicity.

3 Material & Methods

All chemicals, unless stated differently, were ordered in their highest purity from Abcam (Cambridge, UK), Agilent (Santa Clara, California, USA), Beckman Coulter (Brea, California, USA), Bio-Rad Laboratories (Hercules, California, USA), Biozym Scientific (Hessisch Oldendorf, Germany), Carl Roth (Karlsruhe, Germany), Macherey-Nagel (Düren, Germany), Merck KGaA (Darmstadt, Germany), New England Biolabs (Ipswich, Massachusetts, USA), Omnilab (Bremen, Germany), Promega (Madison, Wisconsin, USA), Qiagen (Venlo, The Netherlands), Sigma Aldrich (St. Louis, Missouri, USA), Thermo Fisher Scientific (Waltham, Massachusetts, USA), Th. Geyer (Höxter, Germany) or VWR International (Radnor, Pennsylvania, USA).

3.1 Buffers & Solutions

Unless stated differently, solutions were prepared with ultra-pure water. Solubilized antibiotics were sterile filtered (0.22 µM) unless dissolved in ethanol. Cell culture media were autoclaved or sterile filtered (0.22 µM) if heat labile compounds were present. Protein buffers were filtered (0.45 µM) and degassed, buffers for MALS experiments and heat labile compounds were sterile filtered (0.22 µM).

Antibiotics (1000X)

Ampicillin	100 mg/mL
Chloramphenicol	34 mg/mL in 70% EtOH
Kanamycin	30 mg/mL

Special cell culture media

CDMM	Stock solutions (for 1 l)	Final concentration
	100 mL of 10X Salt mix	
	50 g/L Na_2HPO_4	35.2 mM Na_2HPO_4
	20 g/L NaH_2PO_4	16.7 mM NaH_2PO_4
	9 g/L KH_2PO_4	6.6 mM KH_2PO_4
	9 g/L NaCl	15.4 mM NaCl
	pH 7.0	pH 7.0

200 mL of 5X Casamino acid mix

50 g/L Casamino acids	5 g/L Casamino acids
2.5 g/L L-L-cystein	2.0 mM L-L-cystein
0.5 g/L L-tryptophane	0.2 mM L-tryptophane
0.5 g/L L-proline	0.4 mM L-proline
	sterile filtered

10 mL of 100X Trace elements

4 g/L $(NH_4)_2SO_4$	3.0 mM $(NH_4)_2SO_4$
2.6 g/L $CaCl_2$ – 2 H_2O	2.0 mM $CaCl_2$ – 2 H_2O
2 g/L $MgCl_2$ - 6 H_2O	1.0 mM $MgCl_2$ - 6 H_2O
1 g/L $MnCl_2$ - 4 H_2O	0.5 mM $MnCl_2$ - 4 H_2O
0.1 g/L $CoCl_2$ - 6 H_2O	42 µM $CoCl_2$ - 6 H_2O
0.015 g/L $NaHSeO_3$	9.9 µM $NaHSeO_3$

2 mL of 500X Ironsulfate

2 g/L $FeSO_4$ – 7 H_2O	0.7 mM $FeSO_4$ – 7 H_2O

5 mL of 200X Vitamins

0.06 g/L D-biotin	245 µM D-biotin
0.2 g/L Ca-D-pantothenate	91 µM Ca-D-pantothenate
0.2 g/L pyridoxin	118 µM pyridoxin
	sterile filtered

10 mL of 20% Glucose

200 g/L glucose	11.1 mM (0.2%) glucose, sterile filtered

M9MM	Stock solutions (for 1 l)	Final concentration
	100 mL 10X M9 Salt Solution	
	75.2 g/L Na_2HPO_4 – 2 H_2O	4.21 mMol/L Na_2HPO_4-2 H_2O
	30 g/L KH_2PO_4	22.04 mMol/L KH_2PO_4
	5 g/L NaCl	8.55 mMol/L NaCl
	pH 7.2	pH 7.2
	10 mL 100X Trace Elements	
	5 g/L EDTA	0.134 mM EDTA
	0.83 g/L $FeCl_3$ - 6 H_2O	0.031 mM $FeCl_3$ – 6 H_2O
	84 mg/L $ZnCl_2$	6.2 µM $ZnCl_2$
	13 mg/L $CuCl_2$ – 2 H_2O	0.76 µM $CuCl_2$ – 2 H_2O
	10 mg/L $CoCl_2$ – 6 H_2O	0.42 µM $CoCl_2$ – 6 H_2O
	10 mg/L H_3BO_3	1.62 µM H_3BO_3
	1.6 mg/L $MnCl_2$ – 4 H_2O	0.081 µM $MnCl_2$ – 4 H_2O
	pH 7.5	pH 7.5
	Additives	
	1 g NH_4Cl	1 g/L NH_4Cl, sterile filtered
	20 mL 20% Glucose	22.2 mM (0.4%) Glucose, sterile filtered
	0.5 mL 2M $MgSO_4$	1 mM $MgSO_4$
	0.3 mL 1M $CaCl_2$	0.3 mM $CaCl_2$
	1.0 mL 1 mg/mL biotin	1 mg/L biotin, sterile filtered
	1.0 mL 1 mg/mL thiamin	1 mg/L thiamin, sterile filtered

Common cell culture media

LB medium	10 g/L tryptone, 5 g/L yeast extract, 7 g/L NaCl
LB agar	10 g/L tryptone, 5 g/L yeast extract. 7 g/L NaCl, 16 g/L agar
SOB medium	20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, adjust with NaOH to pH 7.5
SOC medium	20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose, adjust with NaOH to pH 7.5
TB medium	12 g/L tryptone, 24 g/L yeast extract, 4 % (v/v) glycerol
10X TB buffer	23.13 g/L KH_2PO_4, 125.41 g/L K_2HPO_4
IPTG stock	1 M IPTG

Protein purification buffers

DNase stock solution	1 mg/mL DNase I
PrdA lysis buffer	50 mM HEPES pH 8, 500 mM NaCl, 10 mM Imidazole, 5 mM BME, DNase I, cOmplete™ EDTA-free Protease Inhibitor Cocktail
PrdA wash buffer	50 mM HEPES pH 8, 200 mM NaCl, 10 mM imidazole, 5 mM BME
PrdA Ni-NTA elution buffer	50 mM HEPES pH 8, 200 mM NaCl, 5 mM BME, 300 mM imidazole
PrdA SEC buffer	20 mM HEPES pH 8, 200 mM NaCl, 1 mM DTT
PrdB lysis buffer	DNase I, cOmplete™ EDTA-free Protease Inhibitor Cocktail
PrdB wash buffer	50 mM HEPES pH 8, 200 mM NaCl, 5 mM BME
PrdB Ni-NTA elution buffer	50 mM HEPES pH 8, 200 mM NaCl, 5 mM BME, 300 mM imidazole
PrdB Strep elution buffer	20 mM HEPES pH 8, 200 mM NaCl, 1 mM DTT, 8 mM d-desthiobiotin
PrdB storage buffer	20 mM KH_2PO_4/K_2HPO_4 pH 8, 100 mM NaCl
PR lysis buffer	50 mM HEPES pH 8.0, 1 mM DTT, 1 mg/mL lysozyme, DNase I, cOmplete™ EDTA-free Protease Inhibitor Cocktail
PR precipitation buffer	50 mM HEPES pH 8.0, 1 mM DTT, 2 M $(NH_4)_2SO_4$
PR HIC buffer A	50 mM HEPES pH 8.0, 1 mM DTT, 5% (v/v) ethyleneglycol
PR HIC buffer B	50 mM HEPES pH 8.0, 1 mM DTT, 1 M $(NH_4)_2SO_4$

PR IEX buffer A	50 mM HEPES pH 8.0, 1 mM DTT
PR IEX buffer B	50 mM HEPES pH 8.0, 1 mM DTT, 1 M KCl
PR SEC-MALS buffer	50 mM HEPES pH 8.0, 1 mM DTT, 150 mM KCl
PlaA SEC buffer	20 mM Tris pH 8, 150 mM NaCl, 1 mM DTT
PlaB lysis buffer	50 mM HEPES pH 8, 200 mM NaCl, 1 mM DTT, DNase I, cOmplete™ EDTA-free Protease Inhibitor Cocktail
PlaB wash buffer	50 mM HEPES pH 8, 200 mM NaCl, 1 mM DTT
PlaB Strep elution buffer	50 mM HEPES pH 8, 200 mM NaCl, 1 mM DTT, 8 mM d-desthiobiotin
PlaB SEC buffer	20 mM HEPES pH 8, 150 mM NaCl, 1 mM DTT

Bioanalytical application buffers

Western Blot buffer	25 mM TRIS pH 8.3, 152 mM glycine, 15 % (v/v) MeOH
8X SDS PAGE loading buffer, reducing	10 mM Tris/HCl pH 6.8, 3 mM bromophenol blue, 7 % (v/v) SDS, 20 % (v/v) glycerol, 3.5 % (v/v) β-mercaptoethanol
50X TAE buffer	2 M Tris/base, 1 M acetic acid, 50 mM EDTA, adjust to pH 8.5
TEV storage buffer	50 mM Tris/HCl, pH 7.5, 1 mM EDTA, 5 mM DTT, 50% (v/v) glycerol
TFB1	30 mM K-acetate, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15 % (v/v) glycerol, pH 5.8
TFB2	10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% (v/v) glycerol, pH 6.5

3.2 Bacterial strains

All *E. coli* strains were stored as glycerol stocks at -80°C. *Clostridiodes difficile* 630Δerm was cultured at the BRICS, TU Braunschweig.

Table 3.1: *E. coli* strains used in this thesis. The different strains were used for DNA production (marked with *) or protein expression (marked with **)

E. coli strain	Genotype	Reference
DH5α *	F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ^-	(Grant et al., 1990)
BL21(DE3) **	*E. coli* str. B F^- *ompT gal dcm lon hsdS*$_B$($r_B^- m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K-12}$(λ^S)	(Studier et al., 1986)
BL21(DE3) pLysS **	F–, *omp*T, *hsd*S$_B$ (r_B–, m_B–), *dcm*, *gal*, λ(DE3), pLysS, Cm^r.	(Davanloo et al., 1984)
Top10 *	F– mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15ΔlacX74 recA1 araD139 Δ(ara leu) 7697 galU galKrpsL (StrR) endA1 nupG	(Durfee et al., 2008)
XL21 blue *	F´ proAB lacIqZΔM15 Tn10 Tetr/ recA1 endA1 gyrA96thi-1 hsdR17 supE44 relA1 lac	Strategene (San Diego, California, USA)
OmniMAX2 *	F´ {proAB lacIq lacZΔM15 Tn10(TetR) Δ(ccdAB)} mcrA Δ(mrr hsdRMS-mcrBC) Φ80(lacZ)ΔM15 Δ(lacZYA-argF)U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD	Thermo Fisher Scientific (Waltham, Massachusetts, USA)

3.3 Kits and Screens

Kits for molecular biology purposes, listed below, were used following the corresponding manual.

Limited proteolysis	Hampton Research, Aliso Viejo, California, USA
QIAprep MiniPrep	QIAGEN, Venlo, The Netherlands
QIAquick PCR Purification Kit	QIAGEN, Venlo, The Netherlands
QIAquick Gel Extraction Kit	QIAGEN, Venlo, The Netherlands
RUBIC additive screen	Molecular Dimensions Limited, Sheffield, Great Britain Limited, Sheffield, Great Britain

Single conditions from crystallization screens were mixed 1:1 with protein solution in a sitting drop vapor diffusion experiment. Additive screens were mixed in a 1:10 ratio with a stock of a prepared reservoir solution prior to mixing with protein solution. Detailed description of protein crystallization is given in chapter 3.7.1. Crystallization screens used in the presented projects are listed below.

Additive Screen HT	Hampton Research, Aliso Viejo, California, USA
ComPAS	QIAGEN, Venlo, The Netherlands
Index HT	Hampton Research, Aliso Viejo, California, USA
JCSG Core Suite I - IV	QIAGEN, Venlo, The Netherlands
JCSG+	QIAGEN, Venlo, The Netherlands
Mb Class II	QIAGEN, Venlo, The Netherlands
MIDAS™ HT-96	Molecular Dimensions Limited, Sheffield, Great Britain
Morpheus® HT-96	Molecular Dimensions Limited, Sheffield, Great Britain
Morpheus® II	Molecular Dimensions Limited, Sheffield, Great Britain
PACT Suite	QIAGEN, Venlo, The Netherlands
PGA-LM HT-96	Molecular Dimensions Limited, Sheffield, Great Britain
Protein Complex Suite	QIAGEN, Venlo, The Netherlands
Silver Bullets	Hampton Research, Aliso Viejo, California, USA

3.4 General Methods in Molecular Biology

Genes were cloned from *Clostridioides diffcile* 630Δerm genomic DNA (gDNA) provided by Dr. Meina Neumann-Schaal from the DSMZ.

Antibiotics used in culture media or agarose plates were Ampicillin (Amp) at 100 µg/mL, Kanamycin (Kan) at 30 µg/mL, and Chloramphenicol (Cm) at 34 µg/mL.

3.4.1 Generation of chemically competent *E. coli* cells

Rubidium-competent cells were generated in 250 mL batches. Therefore, 250 mL of LB medium was inoculated with the *E. coli* strain desired and grown at 37°C and 130 rpm to an OD_{600} = 0.5. After reaching the OD, the culture was cooled down on ice for 10 minutes and spun down at 4°C, 3000 rpm for 5 minutes. Pelleted cells were resuspended in 100 mL of cold TFB1 and chilled on ice for 5 minutes. Cells were spun down at 4°C, 3000 rpm for 5 minutes and resuspended in 10 mL of cold TFB2. The cell suspension was aliquoted to 50 µL in pre-cooled reaction tubes and immediately flash frozen in LN_2. Aliquots were stored at -80°C until further use.

3.4.2 Agarose gel electrophoresis

In order to verify PCR amplification, a DNA agarose gel electrophoreses was carried out. In detail, 2.4 g of agarose were added to 300 mL of 1X TAE buffer, pH 8 and heated until the agarose was dissolved to yield a 0.8% agarose TAE solution. The solution was cooled down to approximately 60°C. For DNA detection 10 µL Roti®-Safe GelStain (Carl Roth GmbH + Co.KG, Karlsruhe, Germany) was spread drop-wisely in the mini-gel caster (Bio-Rad Laboratories, Hercules, California, USA) and the warm agarose TAE solution was added to the caster to a height of approximately 0.7 cm while Roti®-Safe GelStain was distributed equally in the gel. After letting the gel cool down, it was transferred to a Mini-Sub Cell GT Horizontal Electrophoresis System, 7 x 7 cm (Bio-Rad Laboratories, Hercules, California, USA) and covered with TAE buffer. The comb was removed and up to 50 µL of sample could be applied. DNA markers used were SmartLadder (Eurogentec, Lüttich, Belgium) and GeneRuler™ DNA Ladder Mix (Thermo Fisher Scientific, Waltham, Massachusetts, USA), both applied with 5 µL. The gel was run for 30 minutes, applying 100 V.

3.4.3 Primer Design and PCR

Primer design was performed following a manual approach. Forward primers were determined by 5’ gene boundaries of interest. Melting temperature was adjusted by summing up AT pairs with 2°C and GC pairs with 4°C per base up to 66°C - 68°C for the coding area of the primer. A palindromic cleavage sequence was added at the 5’ end of the sequence depending on the desired restriction enzyme, together with an overhang to enhance cleavage efficiency. Reverse primers were generated equivalently, but cleavage site and overhang were added at the 3’ end before transforming the sequence to the reverse complement. Primers used during the project are listed in table S1. For PCR reactions, the single components were mixed and transferred to a PCR reaction tube as shown in table 3.2.

Table 3.2: PCR ingredients. The PCR gene amplification was performed applying the shown pipetting scheme.

Component	**Volume for**		c_{final}
	25 µL reaction	**50 µL reaction**	
2X Phusion® High-Fidelity PCR Master Mix	12.5 µL	25 µL	
10 µM Primer fw	1.0 µL	2.0 µL	0.4 µM
10 µL Primer rv	1.0 µL	2.0 µL	0.4 µM
DNA template	variable	variable	variable
ddH_2O	ad 25 µL	ad 50 µL	

PCR reactions for gDNA templates needed to be optimized with respect to the amount of introduced gDNA. Best results were achieved by using 1 µL of a gDNA stock with 10 ng/µL for a 25 µL reaction (final concentration 0.4 ng/µL). For plasmid DNA the final concentration varied between 28 ng/µL and 8 ng/µL. The PCR reaction was carried out with a PEQSTAR 2x / Gradient thermocycler (PEQLAB Biotechnologie GmbH, Munich, Germany) with settings shown in table 3.3.

Site directed mutagenesis was performed by polymerase incomplete primer extension (PIPE) cloning, as described in the literature (Ho et al., 1989)

Table 3.3: PCR protocol. Standard protocol for gene amplification from gDNA and plasmids.

Command		Temperature	Time
	Heat lid to	110°C	
	Initial melting	95°C	3 min
Open cycle			
	Denaturation	95°C	30 sec
35x	Annealing	52°C	30 sec
	Elongation	72°C	1 min 30 sec
Close cycle			
	Complementation	72°C	10 min
	Storage	4°C	forever

Certain PCR reactions like polymerase incomplete primer extension (PIPE) cloning, site-directed mutagenesis or those reactions that did not lead to a single amplicon were repeated following a touch-down protocol adopted from the protocol given in table 3.2. Here, the annealing temperature was decreased from 65°C to 53°C within 7 cycles with an increment of 2°C per cycle. Subsequently, 25 cycles with an annealing temperature of 52°C were queued before complementation.

DNA amplicons were either purified by agarose gel electrophoresis as described in chapter 3.4.1 and subsequent gel extraction using the QIAquick Gel Extraction Kit (QIAGEN, Venlo, Netherlands) or were isolated using QIAquick PCR Purification Kit (QIAGEN, Venlo, Netherlands) if PCR performance for a certain reaction was confirmed before. Purified DNA was stored at -20°C until further use.

3.4.4 DNA digestion

Purified DNA was mixed with CutSmart® buffer (New England Biolabs, Ipswich, Massachusetts, USA) and restriction enzymes as shown in table 3.4.

Table 3.4: DNA digestion protocol. DNA digestion was carried out applying the presented pipetting scheme.

Component	**Volume**	**c_{final}**
10X CutSmart® buffer	5 µL	1X
purified DNA	variable	20 ng/µL
Restriction enzyme 1	2.5 µL	variable
Restriction enzyme 2	2.5 µL	variable
ddH_2O	ad 50 µL	

The reaction mixture was incubated at room temperature (RT) for 1 hour. If the digested DNA was plasmid DNA, 1 µL of Antarctic Phosphatase (New England Biolabs, Ipswich, Massachusetts, USA) was added and the reaction was continued at 37°C for 30 minutes. After restriction and eventual dephosphorylation was completed, the reaction mix was heat inactivated at 95°C for 5 minutes. 10 µL of 6X DNA Gel Loading Dye (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was added to the reaction mixture and the sample was analyzed and purified by agarose gel electrophoresis as described in chapter 3.4.1.

3.4.5 DNA ligation

For DNA ligation, the digested insert (i) was added to a digested and dephosphorylated plasmid (p) in 3-fold molar excess. The amount of insert was calculated by formula (1):

$$V_i = 3 * \frac{bp_i * c_p * V_p}{bp_p * c_i} \tag{11}$$

The ligation reaction mix was completed with T4 DNA Ligase reaction buffer and T4 DNA Ligase (New England Biolabs, Ipswich, Massachusetts, USA), as shown in table 3.5.

Table 3.5: DNA ligation. Ligation of digested DNA was carried out applying the presented pipetting scheme.

Component	Volume	c_{final}
10X T4 DNA Ligase reaction buffer	2.0 µL	1X
insert DNA, digested, purified	variable	variable
plasmid DNA, digested, dephosphorylated, purified	variable	variable
T4 DNA ligase	1.0 µL	400 units
ddH_2O	ad 20 µL	

The reaction was carried out for 1 hour at RT. Then, 10 µL were used to start a transformation with competent *E. coli* cells. The remaining reaction mix was transferred to 4°C and stored o/n.

3.4.6 Transformation of DNA to competent *E. coli* cells

To transform DNA into *E. coli* cells, a 50 µL aliquot of chemically competent *E. coli* cells was slowly thawed on ice. 1 to 10 µL of the desired DNA solution was added carefully to the cell suspension, which was incubated on ice for 5 minutes. Subsequently, a heat shock was performed by heating the cells to 42°C for 90 seconds. Afterwards, the cells were incubated on ice for another 5 minutes. Then the cell suspension was supplemented with 1 mL of preconditioned SOC medium and incubated in a Thermomixer® R (Eppendorf, Hamburg, Germany) at 37°C with 800 rpm for 30 – 60 minutes. If plasmid DNA was transformed, 50 µL of the cell suspension were spread on an agar plate containing the desired antibiotics. In case of a transformed ligation batch, the cells were spun down for 3 minutes with 5000 rpm and 20°C, resuspended in 50 µL of SOC medium and spread on an agar plate containing the desired antibiotics. The inoculated plates were incubated o/n at 37°C.

3.4.7 Colony PCR and sequencing

Clones arising from a ligation batch were checked for correct gene insertion by colony PCR and sequencing. To run a colony PCR, the Red-*Taq* DNA-Polymerase-MasterMix (VWR®, Radnor, Pennsylvania, USA) was prepared together with the primers necessary for gene amplification. Usually, primers annealing to the T7 promotor and terminator were used.

Table 3.6: Colony-PCR ingredients. A colony-PCR was prepared applying the shown pipetting scheme.

Component	**Volume for 10 * 50 µL reactions**	**c_{final}**
2X Red-*Taq* DNA-Polymerase-Mastermix	250 µL	1X
Primer fw	10 µL	0.02 pmol/µL
Primer rv	10 µL	0.02 pmol/µL
ddH$_2$O	230 µL	

The PCR reaction mix was transferred to PCR reaction tubes as 50 µL aliquots. Single colonies from o/n agar plates were picked with a sterile toothpick and used for introducing the DNA template to the PCR reaction mix. Further, the toothpick was used to inoculate 5 mL of LB medium supplemented with desired antibiotics, which was then incubated o/n at 37°C and 130 rpm.

The PCR was performed following the protocol described in chapter 3.4.3 and the size of the amplified DNA was documented by agarose gel electrophoresis as described in chapter 3.4.2.

Cultures which showed a correct insert size were harvested and plasmid DNA was isolated using the QIAprep Miniprep-Kit (Venlo, The Netherlands). The isolated DNA was sent for sequencing (Eurofins, Luxemburg). Once the correct gene insertion was confirmed, DNA was transformed to different *E. coli* strains as described in chapter 3.4.6 and after o/n culture, glycerol stocks were generated by mixing 500 µL of culture with 500 µL of 86 % (v/v) glycerol in a cryovial. The cryo-protected suspension was flash frozen in LN$_2$ and stored at -80°C. A small amount of this material was transferred into fresh medium to inoculate another culture.

3.5 Protein Production and Purification

During the working process, protein samples were generally kept on ice or at 4°C. If not used in the near future, purified protein solutions were aliquoted to ≤ 100 µL and were flash frozen in LN2.

3.5.1 Recombinant protein production in *E. coli*

After a plasmid had been successfully transformed into an *E. coli* expression strain as described in chapter 3.4.6, a pre-culture was incubated o/n in corresponding medium with added antibiotics. The OD_{600} was determined using the WPA Biowave CO8000 Cell Density Meter. For the main culture, the desired volume of corresponding medium with added antibiotics was inoculated with the pre-culture to an OD_{600} = 0.05. The culture was incubated at 37°C and stirred with 130 rpm, until an OD_{600} = 0.5 was reached. In the following the temperature was lowered to 20°C and expression was induced by addition of 50 – 1000 µM IPTG. Protein production carried out o/n and the cells were pelleted using a Sorvall™ RC6+ Centrifuge (Thermo Fisher, Waltham, Massachusetts, USA) at 4°C with 6000 rpm for 15 minutes. Subsequently, the cells were transferred into a 50 mL centrifuge tube and flash frozen in LN2. The cells were stored at -20°C until used.

3.5.2 Expression test

To test optimal conditions for expression of certain recombinant proteins, the parameters temperature, IPTG concentration and time were varied to identify their influence on protein expression. To do so, a 200 mL culture of desired medium with antibiotics was inoculated with an OD_{600} = 0.05 from an o/n culture. This main culture was incubated at 37°C and 130 rpm, until the OD_{600} reached a value of 0.5. Then the main culture was split into four 50 mL cultures. Two of the cultures were induced by the addition of 50 µM IPTG, the other two were induced by the addition of 500 µM IPTG from a 1 M IPTG stock solution. From both series, one flask each was transferred to 20°C, the remaining flasks were kept at 37°C. To check on expression of soluble protein, 10 mL of culture were taken from each flask after 3 hours and after o/n expression and were transferred to 15 mL centrifuge tubes. The cells were pelleted by centrifugation at 4°C, 4000 rpm for 10 minutes and the supernatant was discarded. Pelleted cells were flash frozen in LN2.

For protein expression analysis, cell pellets were thawed and resuspended in 1 mL cold standard lysis buffer. The cells were disrupted by sonication with three 30 seconds cycles with a Bandelin Sonopuls, equipped with a MS73 sonotrode, set to 50% cycles and 70% amplitude. In between the sonication steps, the sample was cooled on ice. After lysis the suspension was transferred to a 1.5 mL reaction tube and cell debris was removed by centrifugation at 4°C, 16.000 rpm for 30 minutes. The supernatant was transferred to a fresh reaction tube and affinity beads were added. Depending on the affinity tag fused to the recombinant target protein, certain volumina of 50% bead suspensions were added to the crude lysate. For Strep-tagged protein, 5 µL Strep-Tactin®XT high-capacity beads (IBA Lifescinces, Göttingen, Germany), for His$_6$-tagged protein 7 µL Ni Sepharose 6 Fast Flow (Thermo Fisher Scientific, Waltham, Massachusetts, USA) or for MBP-tagged protein 5 µL Amylose Resin (New England Biolabs, Ipswich, Massachusetts, USA) were added. The mixtures of lysate and affinity beads were incubated at 4°C on a rotating mixer for 1 hour. Afterwards, the resins were pelleted by gentle centrifugation at 4°C, 500 rpm for 1 minute and the supernatant was carefully removed. The remaining beads were washed with fresh standard lysis buffer and spun down three times. Finally, 10 µL of 8X SDS-PAGE loading buffer were added to the bead material. To check for insoluble expression products, a tip of pelleted cell debris was transferred to 50 µL of standard lysis buffer and 10 µL 8X SDS-PAGE loading buffer. The total content of soluble protein was addressed by adding 10 µL 8X SDS-PAGE loading buffer to 50 µL of crude lysate. All SDS-PAGE samples were analyzed by SDS-PAGE as described in chapter 3.5.9.

The gels were inspected with respect to the amount of purified, soluble protein that was present in the bead fraction for each expression condition. The identified condition for highest productivity was upscaled to 1 L batches.

3.5.3 Selenomethionine derivatization

As described in chapter 1.3.1., selenomehtionine (SeMet) labelling of recombinant protein can become necessary if experimental phasing is required to solve a protein structure from an X-ray crystallography experiment. To achieve a high incorporation rate with sufficient amount of protein yield, the production of the target protein in an M9MM, supplemented with SeMet, was optimized. The *E. coli* strain transformed with the plasmid of the target protein was cultured in an o/n pre-culture at 37°C and 130 rpm. Two times 1 L TB medium, supplemented with TB buffer and necessary antibiotics were inoculated with an OD$_{600}$ = 0.05 and also incubated at 37°C and 130 rpm in another o/n culture. This culture was spun down at RT, 6.000 rpm for 15 minutes. The resulting cell pellet was resuspended in 500 mL M9MM and again spun down at RT, 6.000 rpm for 15 minutes to remove residual components of TB medium. The washed cell pellet was then resuspended in two times 1 L M9MM, conditioned to 20°C and was incubated at 20°C, at 130 rpm for 15 minutes. After that adaption time, SeMet was added to a final concentration of 50 mg/L and after another 15 minutes, the protein expression was induced by addition of the desired amount of IPTG. The expression was carried out at 20°C and 130 rpm o/n. Cells were harvested by centrifugation at 4°C, 6.000 rpm for 15 minutes and the cell pellet was transferred to a 50 mL centrifuge tube. Cells were frozen in LN2 and stored at -20°C until further use.

3.5.4 ^{15}N labelling of recombinant proteins

To be able to generate a “protein finger print” by an HSQC-experiment, the target protein needs to be labelled by ^{15}N. The expression needed to be performed in the smallest volume and at highest cell densities possible, due to the high costly incorporation of ^{15}N. Therefore, expression as a 2 L fermenter batch was addressed. To achieve a good yield and high incorporation yield, the expression protocol was optimized together with Daniela Gebauer and Dr. Joop van den Heuvel, both working at RPEX at the HZI.

The Labfors 5 fermenter (Infors HT, Germany) was filled with the heat stable components of the M9MM (100 mL 10X M9 salt solution, 0.5 mL 2M $MgSO_4$, 0.3 mL 1M $CaCl_2$, 10 mL 100X trace elements and water) and was autoclaved fully assembled. After the fermenter cooled down to RT, heat labile compounds were added through a 0.22 µM sterile filter (5 mL 200 mg/mL $^{15}NH_4Cl$, 20 mL 20% glucose, 1 mL 1 mg/mL biotin and 1 mL 1 mg/mL thiamin). The temperature was adjusted to 20°C and the oxygen probe was calibrated. For inoculation, an o/n culture was grown in TB medium, supplemented with TB puffer and desired antibiotics. The cells were pelleted at RT and 6.000 rpm for 15 minutes and resuspended in 500 mL M9MM (without ^{15}N) to remove residual components of the TB medium. The cells were pelleted again and resuspended in 200 mL of M9MM (without ^{15}N) and the OD_{600} was determined. With this inoculum the fermenter was inoculated with an $OD_{600} = 2.0$ and cells were allowed to adapt for 30 minutes. After adaption, protein expression was started by addition of 1 mM IPTG. During expression, the pH was monitored, anti-foam was applied and oxygen was regulated by stirring speed.

After expression, cells were harvested and spun down at 4°C, 6.000 rpm for 15 minutes. Cell pellets were distributed on 50 mL centrifuge tubes and flash frozen in LN2. The frozen cells were stored at -20°C until processed further.

3.5.5 Cell lysis by homogenization

A cell pellet harvested from a 1 L o/n culture as described in chapter 3.5.1, was resuspended in an appropriate volume of cold lysis buffer. The suspension was stirred vigorously in a beaker at 4°C for 10 minutes and afterwards passed through two sheets of filter paper. The cell suspension was then transferred to the inlet vessel of the pre-equilibrated EmulsiFlex-C3 (Avestin Europe GmbH, Mannheim, Germany). The suspension was passed through the homogenizer twice with a pressure of 16.000 psi and was collected in centrifuge tubes. Lysed cells were pelleted by centrifugation at 4°C, 16.000 rpm for 1 hour. The lysate was subjected to downstream processing.

3.5.6 Recombinant protein affinity chromatography techniques

Lysates from recombinant protein expression generated as described in chapters 3.5.1, 3.5.3, 3.5.4 and 3.5.5 were further addressed by the appropriate affinity chromatography technique as a first step of purification. Lysate that was planned to be subjected to Strep-Tactin® material was additionally cleared at 4°C, 100.000 x g for 1 hour.

The affinity columns were attached to ÄKTA™ systems (GE Healthcare, Chicago, Illinois, USA) and equilibrated with desired buffer (for an overview of buffers used, see chapter 4.1). Subsequently the lysate was passed through the column using a super loop or a sample pump, the flow through was collected. Unspecifically bound proteins were removed by washing until UV absorption of eluting solvent reached a constant level. Elution was initiated by addition of competitive compounds, such as imidazole for the Ni-NTA and d-desthiobiotin for the Strep-Tactin® purification. Running conditions for HisTrap™ HP and columns packed with Strep-Tactin® Superflow® high capacity are listed in table 3.7.

Table 3.7: Affinity chromatography columns. Recombinant proteins were tagged either with a His_6- or a Strep-tag. Accordingly, the following materials were used.

	His_6-tag	Strep-tag
Column	5 mL HisTrap™ HP	8 mL Strep-Tactin® Superflow®
Flow rate	3 mL/min	2 mL/min
Pressure limit	0.5 MPa	0.5 MPa
Elution	0 – 300 mM gradient imidazole	7 mM d-desthiobiotin
Avoided substances	DTT, EDTA	SDS, Guanidine/HCl

Eluting material was collected and fractionated. The purification success was monitored by SDS-PAGE with cell pellet, crude lysate, flow through and elution samples. An SDS-PAGE was performed as described in chapter 3.5.9. If applicable, eluted protein was mixed with recombinant TEV protease in a gravimetric ratio of 1:100 (Invitrogen, Carlsbad, California, USA) and dialyzed o/n to remove the affinity tag. TEV digestion control was performed by SDS-PAGE and the cleaved protein was collected and concentrated for further processing. If the protein was not used immediately, it was flash frozen as ≤ 100 µL aliquots in LN_2 and stored at -80°C.

3.5.7 Cultivation of *Clostridioides difficile*

Cultivation of *C. difficile* was performed at the BRICS, TU Braunschweig with the help of Dr. Meina Neumann-Schaal and Dr. Julia Hofmann. All cultures were grown at 37°C in a water bath. Medium used was anaerobic CDMM. To achieve high chances of germination, several pre-cultures of 50 mL were inoculated from a spore stock and incubated o/n. A second pre-culture of two times 100 mL was inoculated with an $OD_{600} = 0.05$ and incubated o/n. The main culture consisted of eight times 0.8 L CDMM supplemented with 0.1 g/L proline, which were inoculated with an $OD_{600} = 0.05$ and incubated until the OD_{600} reached a value of 0.4. This density marks the end of the exponential growth phase, where the highest amount of proline reductase complex was assumed, as described in chapter 1.1.3.. The cells were then harvested by centrifugation at 4°C, 4.000 rpm for 20 minutes. Four pellets were collected, each from 1.6 L culture, and transferred to four 50 mL centrifuge tubes. The pellets were flash frozen in LN2 and stored at -80°C until further use.

3.5.8 Native purification techniques

The purification protocol for isolation of native proline reductase complex from *C. difficile* was adopted from Kabisch et al. (Kabisch et al., 1999) and iteratively optimized until the sample quality was sufficient for reliable biophysical characterization. Decisions towards variations were made based on a reduction of required purification steps, higher selectivity of the affinity material towards the proline reductase complex and improved thermostability of the purified protein and resulted in the following protocol.

One cell pellet from 1.6 L *C. difficile* culture, generated as described in chapter3.5.7, was resuspended in *C. difficile* lysis buffer (see chapter 4.1) to a total volume of 40 mL. The suspension was transferred to a cylindrical Schlenk flask, which was aerosol-tightly attached to a sonicator (Bandelin Sonopuls, equipped with a MS73 sonotrode). The flask was placed in a beaker filled with an ice/water mixture. Sonication was carried out on ice with cycles set to 10% and amplitude to 50% for 30 minutes. The suspension was then transferred to a measuring cylinder and 40 mL of *C. difficile* precipitation buffer (see chapter 4.1) were added slowly while stirring. The suspension was stirred for 5 minutes, transferred to two high performance centrifuge tubes and centrifuged at 4°C with 100.000 xg for 1 hour. Equipment used during cell lysis and precipitation was covered with a sporozide (Incidin™ Active, ECOLAB Healthcare, Mohnheim am Rhein, Germany) for at least 1 hour and autoclaved (except high performance centrifuge tubes).

For identification of most suitable HIC material, the HIC selection kit was used. Equal amounts of crude lysate were applied on 1 mL columns and washed. The elution gradient of KCl was set to 330 mM within 5 mL and the concentration was kept there for 20 mL to remove impurities. Subsequently, the proline reductase complex elution was initiated by increasing the KCL concentration to 400 mM within 10 mL. This gradient could be set even steeper to avoid unnecessary sample dilution. The concentration was kept for 20 mL and fractions were collected. Remaining bound material was eluted by increasing the KCL concentration to 1 M.

When the HiTrap Butyl HP HIC column was identified to be most suited, the supernatant was decanted carefully and applied on 5 mL column material (GE Healthcare, Chicago, Illinois, USA), pre-equilibrated with *C. difficile* HIC buffer B (for *C. difficile* HIC buffer A and B see chapter 4.1). The flow-through was collected. The column was washed with *C. difficile* HIC buffer B and elution of bound material was achieved by application of a negative linear gradient from 100% *C. difficile* HIC buffer B to 0% over 24 mL. The elution volume was collected and fractionated in 1 mL samples, which were analyzed by

SDS-PAGE as described in chapter 3.5.9. The gel was inspected for the distinctive pattern of the proline reductase complex subunits PrdAβ (45 kDa) and the double band for PrdAα (22 kDa) and PrdB (25 kDa). Fractions containing this pattern were pooled, despite remaining impurities.

The pooled fractions were applied on an 8 mL MonoQ IEX column (GE Healthcare, Chicago, Illinois, USA), pre-equilibrated with *C. difficile* IEX buffer (5% *C. difficile* IEX buffer B. For *C. difficile* IEX buffer A and B see chapter 4.1). The flow-through was collected and the column was washed with *C. difficile* IEX buffer (5% *C. difficile* IEX buffer B). Elution of bound material was performed by gradually increasing the KCl concentration, as shown in table 3.8.

Table 3.8: IEX MonoQ elution gradient. The purification of HIC-purified proline reductase was performed with the shown gradient. The column was equilibrated with 5% buffer B. IEX buffer A was gradually replaced by IEX buffer B.

	from %B	to %B	Volume [mL]
Gradient	5	35	16
Plateau	35	35	8
Gradient	35	50	2
Plateau	50	50	10
Gradient	50	100	2

Samples were fractionated in 1 mL and analyzed on SDS-PAGE as described in chapter 3.5.9. Again, the gel was inspected for appearance of the distinctive pattern for the proline reductase subunits. The proline reductase complex eluted between 350 and 500 mM KCl and was collected. The solution was either aliquoted to 100 µL and flash frozen in LN2 or immediately concentrated with a 2 mL protein concentrator with 10 kDa MWCO (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The solution was concentrated to approximately 500 µL and applied to a Superose6 10/30 increase column (GE Healthcare, Chicago, Illinois, USA), pre-equilibrated with *C. difficile* SEC-MALS buffer (see chapter 4.1) and attached to a SEC-MALS system (Wyatt/Agilent). Fractions were analyzed by SDS-PAGE as described in chapter 3.5.9., pooled and concentrated to 10 mg/mL. If subsequent TEM was addressed, the proline reductase complex was crosslinked by the addition of DSSO, dissolved in DMSO, at a maximum concentration of 15 mM. The reactions were carried out for 30 minutes in the dark at RT and were stopped by the addition of 2 M Tris buffer at pH 8.0. Pipetting schemes are provided in table 3.9.

Table 3.9: Proline Reductase DSSO crosslinking protocol. Crosslinking of the purified proline reductase complex was achieved by adding DSSO to the protein solution. The pipetting scheme for different final concentrations is shown. Protein solution was present at 10 mg/mL. DSSO was dissolved in DMSO with a stock concentration of 50 mM.

$V_{protein}$ [µl]	V_{DSSO} [µl]	V_{buffer} [µl]	V_{Tris} [µl]	c_{DSSO} [mM]
32.7	0.7	-	5.0	1
15.0	1.6	23.4	10.0	2
15.0	2.4	21.8	10.0	3
15.0	3.2	20.2	10.0	4
15.0	6.4	18.6	10.0	8
15.0	8.0	17.0	10.0	10
15.0	12.0	13.0	10.0	15

The crosslinked protein complex was applied to SEC-MALS, once again, and fractions were collected for further analysis. The proline reductase complex containing fractions were pooled and either immediately used or flash frozen as 100 µL aliquots in LN2.

3.5.9 SDS-PAGE

For SDS-PAGE, a desired volume of sample was mixed with 8X SDS PAGE loading buffer and heated to 95°C for 5 minutes. A Mini-PROTEAN® Tetra System (Bio-Rad Laboratories, Hercules, California, USA) was assembled with Mini-PROTEAN® TGX Gels (Bio-Rad Laboratories, Hercules, California, USA) and filled with Rotiphorese® SDS-PAGE buffer at 1X concentration (Carl Roth, Karlsruhe, Germany). As protein standard, 5 µL Unstained Protein Molecular Weight Marker (Fermentas, Waltham, Massachusetts, USA) were applied to one chamber. The volume of sample varied between 2 and 15 µL. Running conditions were 200 V for 30 minutes. The finished gel was removed from the chamber and rinsed with water. Staining was performed with InstantBlue™ (Merck KGaA, Darmstadt, Germany) for 30 minutes. Afterwards the gel was transferred to water to reduce background staining.

3.5.10 Blue native PAGE

Blue native page was carried out using a XCell SureLock™ Electrophoresis Cell and a NativePAGE™ 4-16% BisTris Gel. As anode buffer 20X Native PAGE™ Running Buffer was diluted to 1X concentration and filled into the outer electrophoresis chamber. For the cathode buffer, 1.5 µL of NativePAGE™ Cathode Buffer Additive was added to 300 µL of 1X anode buffer, which was then filled to the inner electrophoresis chamber.

The marker was constituted by dissolving the content of one vial HMW calibration kit for electrophoresis (Amersham, GE Healthcare, Chicago, Illinois, USA) in 100 µL of NativePAGE™ Sample Buffer and 5 µL of that marker were applied onto the gel. Also 12 µL of sample were mixed with 4 µL NativePAGE™ Sample Buffer and 15 µL of the sample were loaded into the gel chambers. The voltage for electrophoresis was set to 10 V for 30 minutes and was subsequently increased to 100 V for 3 hours. The finished gel was removed from the chamber and rinsed with water. Staining was performed with InstantBlue™ (Merck KGaA, Darmstadt, Germany) for 30 minutes. Afterwards the gel was transferred to water to reduce background staining (all products of Invitrogen/Life Technologies, Carlsbad, California, USA).

3.5.11 Limited Proteolysis

To identify protease-resistance domains, 12 x 175 µg of PrdA were each mixed with 14 µL of 1:100 diluted proteases provided by the ProtiAce-Kit and ProtiAce-Kit2 (Hampton Research, Aliso Viejo, California, USA). The mixtures were incubated at 37°C and samples were taken at different time points. 8x SDS loading buffer was added and samples were heated to stop proteolysis. The effect of each protease was monitored by applying the samples on SDS-PAGE. The proteases were compared in their specific digestion behavior on PrdA by visual interpretation of the band pattern generated on the SDS-PAGE.

3.5.12 Semi-dry Western blot for N-terminal sequencing

For Semi Dry Western Blots, an SDS-PAGE was prepared as described in chapter 3.5.9, but PageRuler™ Prestained Protein Ladder (Thermo Fisher, Waltham, Massachusetts, USA) was used as marker and as a positive transfer control. The sample pattern was loaded twice. The stacking gel was removed and the piece for western blotting was cut in size. The double sample pattern was split in two halves, wherefrom one was immediately stained with InstantBlue™ (Merck KGaA, Darmstadt, Germany) for control purposes. Four pieces of Whatman filter paper and PVDF membrane were cut to pieces with sizes according to the gel. The filter paper was soaked with Western Blot buffer. The PVDF membrane was first activated in MeOH and then transferred to water for two minutes. Then it was equilibrated in Western Blot buffer for 5 minutes and placed on top of two layers of filter papers on the anode of the Trans-Blot Turbo System (Bio-Rad Laboratories, Hercules, California, USA). The gel was positioned on the membrane and covered with two further sheets of soaked filter papers. Any air bubbles were removed. The cathode was put on top of the stack and closed such that the cassette was transferable to the blotting system. The voltage was set to 25 V and the blotting time was set to 30 minutes. The blotting efficiency could be estimated by the transfer of the pre-stained protein marker. The SDS-PAGE was further stained with InstantBlue™ (Merck KGaA, Darmstadt, Germany) to check for non-transferred proteins. For isolation of membrane parts that carry the unstained target protein, the stained SDS-PAGE that was separated and stained before served as a template and corresponding spots were cut from the membrane. The pieces of membrane were washed in ethanol/acetic acid and subjected to N-terminal sequencing.

3.5.13 N-terminal sequencing

N-terminal sequencing was performed by Beate Jaschok-Kentner at the Helmholtz Centre for Infection Research (Braunschweig, Germany). Protein samples were immobilized on a PVDF membrane by Western blotting. The PVDF membrane was carefully washed in ethanol/acetic acid. The cut-out bands were analyzed using an ABI 494 protein sequencer. The N-terminal residues were derivatized with PITC in a basic environment and cleaved after a pH-shift to acidic environment. Cleaved N-terminal residues were then analyzed using a reverse phase chromatography. The derivatization and cleavage steps were repeated 10 to 15 times. Identification of separated amino acids was achieved by comparison of the elution volume with standards.

3.6 Biophysical methods, principles and procedures

The following section provides basic explanations of biophysical principles and experimental procedures for the set of biophysical methods that have been used in this thesis.

3.6.1 Microscale Thermophoresis

The physical phenomenon of thermophoresis describes the movement of a molecule in a medium due to a local temperature gradient. Most often molecules move from a higher temperature towards a lower temperature. The physical principles of this phenomenon are not fully understood yet, but its application on proteins has been developed into method to monitor interactions. Depending on the detection system, proteins either need to be labelled with a fluorophore or can be detected using the intrinsic fluorescence of tryptophane residues. An IR laser induces a sharp locally and temporally defined temperature gradient in a capillary filled with protein solution. This gradient induces thermophoresis of the protein molecules and within seconds the formerly equally distributed protein molecules will find a new distribution state with less molecules in the heat center, which leads to a lower detection signal. Turning off the IR laser lets the capillary cool down in a defined speed, which leads to molecular motion back to an equal distribution and to the initial signal intensity. During the whole experiment, speed and end states of the new equilibrium depend on many factors, mainly the hydration shell, protein charge, size and shape among others. Applying this technique to a constant concentration of a molecule with increasing concentration of a ligand, the affinity of both molecules to each other can be determined (Jerabek-Willemsen et al., 2014).

Experimentally, the protein was labelled with a Cy5-NHS reagent, active against lysines, by the addition of 30 µM dye to 10 µM protein solution in an amine free buffer in a total volume of 200 µL. The reaction was carried out in the dark at RT for 30 minutes. Removal of aggregated sample material and free label was performed by a buffer exchange with a NAP-5 column (GE Healthcare, Chicago, Illinois, USA), equilibrated with protein buffer. Eluting material was collected, fractionated and checked by UV-Vis-spectrometry for protein content. First tests for an MST setup included capillary pre-tests in which ideal intensity curves and sufficient protein concentrations were targeted. Improvements can be achieved by choosing appropriate coating (standard, hydrophobic or premium coating), adjusting the LED power for excitation or dilution of the sample. The ideal signal has 200 to 5000 counts and does not show any shoulders. Once those pre-tests were accomplished, a 1:1 dilution series with 16 concentrations of the ligand in the corresponding protein buffer was set up. For each concentration, 10 µL of protein and 10 µL of ligand were mixed and incubated for 10 minutes. Capillaries were filled with the samples and measured with the Monolith™ NT.115 (Nanotemper, Munich, Germany). A cap-scan was collected for verification of equal sample distribution and signal intensity, which was followed by the thermophoresis experiment (Jerabek-Willemsen et al., 2014).

3.6.2 Thermal shift assay

In a thermal shift assay, the melting temperature of a purified protein sample is determined. This is achieved by adding SYPRO® orange protein stain to the protein in a screen of different buffers or additives and by gradually heating the sample from 4 to 95°C in a real-time PCR cycle system. Upon denaturation, hydrophobic moieties of the protein core will become solvent accessible. This leads to an interaction of the added dye with the protein, enhancing the fluorescence signal. Generating the first derivative of dF/dT allows determination of the melting temperature (Wienken et al., 2010; Jerabek-Willemsen et al., 2014).

This technique is useful for proteins that have a low stability. Buffer parameters such as pH and salt concentration were generally checked for any protein purified. Further, additive screening allows identification of compounds that stabilize the target due to binding or shielding actions. In addition, these results can potentially give more information with respect to domain architecture or oligomerization if double-peaks or non-symmetric peaks occur (Boivin et al., 2013; Chari et al., 2015).

It has been shown that an increased melting temperature and melting behavior correlates with enhanced target crystallization (Boivin et al., 2013; Chari et al., 2015). This link can be explained by reduction of entropy upon binding between a protein and its ligand. Usually, interaction sites are flexible regions that take a defined and stiffer shape than before when a ligand is bound to this specific area. As a result, the protein-ligand complex has reduced flexibility and higher rigidity and it is more likely to be crystallized successfully. The same principles can be applied to any other technique in the field of structural biology.

The experiments were performed in a 96-well format. First, a suitable protein concentration was determined, usually between 0.5 mg/mL and 2 mg/mL (0.05 mg/mL to 0.2 mg/mL final concentration). A screen was set up by mixing 5 µL protein solution, 5 µL SYPRO® orange protein stain and 40 µL buffer. Alternatively, the buffer fraction was spilt to 25 µL of an additive screen and 25 µL 2X buffer. A qPCR cycler (Bio-Rad Laboratories Hercules, California, USA) was used to gradually heat the samples and to detect the fluorescence signal.

3.6.3 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy can provide information on secondary structure (SS) elements in proteins, as circularly polarized light gives distinct signals for α-helices and β-sheets. Circularly polarized light is defined as wave of light with a rotating amplitude vector when the wave is projected on a plain. It can be described as a superimposition of two linearly polarized light waves with the same amplitude heights that lie perpendicular to each other and have a phase shift of 90°. Circularly polarized light changes its polarization upon interaction with chiral molecules. Those molecules absorb linearly polarized light in dependency of the angle of the incoming wave, relative to the orientation of the sample. As a result, the circularly polarized light is more strongly weakened in the amplitude of one linearly polarized wave component than in the other, and the superimposed wave is then elliptically polarized. The ellipticity is observed and monitored over a range of energies. SS elements of proteins generate distinct minima in this spectrum in the range of 190 – 240 nm. At 208 and 220 nm, α-helices generate characteristic double minima. β-strands are identified by a minimum at 215 nm. The range of 240 – 320 nm can give information on aromatic residues. The aromatic side chains are not chiral by themselves, but will be detectable by CD spectroscopy as soon as they are embedded in chiral environment such as a protein core (Greenfield, 2009).

The measurements reported here were performed with a Jasco-815 CD Spectrometer (Jasco Deutschland GmbH, Pfungstadt, Germany). The protein sample was dialyzed o/n against a 20 mM potassium phosphate buffer with a KCl concentration as low as possible. Cuvettes were washed at least five times with MilliQ water and dried. The blank measurement was performed with 200 µL of dialysis buffer. The protein concentration was adjusted with dialysis buffer. Then the protein sample was measured with the same settings as the blank. SS estimation was performed by the online tool K2D3 (Louis-Jeune et al., 2012).

3.6.4 *In-situ* dynamic light scattering

An *in-situ* dynamic light scattering (DLS) experiment was performed together with Dr. Arne Meyer and the company Xtal-concepts (Hamburg, Germany). The SpectroLight 600 spectrometer provides the possibility to detect time-resolved light scattering in a protein solution with volumes as small as 100 nL in a 96-well format. A coherent and monochromatic laser is used to generate Rayleigh-scattering from the particles in solution. The scattered light waves interfere with each other and show fluctuation, caused by molecular movement of the particles in solution. From this interference pattern, the movement velocity can be determined, which is dependent on the hydrodynamic radius of a particle. This enables a direct measurement of particle size distribution as a high throughput screening paired with very little sample consumption.

The protein solution was mixed in a PCR reaction tube with buffers and additives and placed as a 1 µL droplet under oil.

3.6.5 Multi angle light scattering

Multi angle light scattering (MALS) is used for the accurate determination of MW and particle size. Both values arise from specific scattering events (Patel et al., 2018).

First, the overall intensity of scattered light I_s depends on a constant, specific for the experimental setup and the sample, its concentration c and its molar mass M.

$$I_s = const. * c * M \tag{11}$$

The setup specific constant can be explained in more detail by

$$I_s = K * I_i * \frac{V_s}{l^2} * c * M \tag{12}$$

where K is the scattering constant, I_i is the intensity of the incoming laser, V_s is the scattering volume and l the distance between scattering sample and the detector. In this equation, except for K all parameters are known. K can be resolved as

$$K = \frac{(2\pi n_0)^2 \left({dn}/{dc}\right)^2}{N_A \lambda^4} \tag{13}$$

where n_0 provides the refractive index of the solvent, dn/dc is the refractive index (RI), N_A the Avogadro number and λ the wavelength of incoming light. For regular proteins, the RI is considered as a fixed value of 0.185 mL/g. The value for RI of the sample is directly proportional to its concentration. Therefore, measuring the RI with a differential refractometer provides a robust method to determine a protein concentration. As dn/dc is known and the concentration is determined, the scatter intensity is directly proportional to the MW of the protein (Wyatt, 1998).

On the other hand, the principle of size determination is based on the anisotropic scattering, which occurs as soon as the scattering particle exceeds a radius of 10 to 15 nm, which is true for many biomolecules. For particles larger than 15 nm, anisotropic scattering occurs, which can be detected with a multi angle detector setup. The measured intensity of low angle scattered light correlates to Rg.

The measurements were performed on an Agilent chromatography system, equipped with WYATT detection systems (Optilab® T-rEX refractometer, MiniDawn TREOS II MALS detector).

The samples were purified by affinity chromatography and were tested for stability and solubility beforehand. Stable and soluble samples of sufficient purity were concentrated. The sample volume may not exceed 100 µL. The amount of material needed for a reliable measurement varies with protein size. For smaller proteins a minimum mass of 200 µg was aspired. The amount needed decreases with increasing protein size. A SEC column was chosen with respect to its separation range and was equilibrated with running buffer with at least 2 column volumes (CV). Then the experiment was queued, monitored and analyzed by ASTRA® software (Wyatt Technology Corporation, USA).

3.7 Methods in structural biology

Strengths and weaknesses of different methods in structural biology were described in chapter 1.3. Here, protocols followed are described to conclude structural information from adequate experiments.

3.7.1 X-ray crystallography

The principles of X-ray crystallography are explained in more detail in chapter 1.3.1. Samples for protein crystallization required an excellent purity and were analyzed in terms of homogeneity and stability, as well.

3.7.1.1 Protein crystallization

In general, protein crystals were grown in vapor diffusion experiments. Regular screening was performed with the Honeybee 961 pipetting robot (Digilab®, Hopkinton, Massachusetts, USA), which transferred 60 µL of reservoir solution from a deep-well storage block to the reservoir chamber of an 96-3 Intelli-plate® (Art Robbins Instruments, Sunnyvale, California, USA) and dispensed 200 nL each of reservoir and protein solution into three sitting drop wells. The chambers were sealed and stored in a Rock Imager®

automated microscope (Formulatrix, Bedford, Massachusetts, USA), monitored and inspected over a period of 48 days.

Additive screening was performed by prefilling the reservoir chamber manually with 96 times 6 µL additive solution and 54 µL of a defined reservoir. The prefilled Intelli-Plate® was then further processed with the Honeybee.

For seeding, a seed stock was prepared by crushing a single crystal in 50 µL of its reservoir solution. The suspension was further diluted when necessary. Crystallization plates were set up using the OryxNano pipetting robot (Douglas Instruments, East Garston, UK) in order to mix the protein solution with both, the reservoir and the seed stock.

Detected crystals were fished using a micro-mesh or loop, attached to a pin holder, fixed to a magnetic cryo wand. If necessary, the crystals were transferred to a soaking solution. Before flash-freezing the crystals in LN2, the crystals were transferred to a reservoir solution, supplemented with sufficient amount of cryoprotectants, i.e., 10% (v/v) glycerol.

3.7.1.2 Data collection, processing and interpretation

Crystals were tested for diffraction using an in-house X-ray generator equipped with a Cu K-α rotating anode. Full and high-quality datasets were collected at the beamlines of various synchrotrons. SAD experiments were in general performed at a synchrotron source with a tunable wavelength. For SAD phasing, highly redundant data were collected with a relatively low dose per image. Data indexing, integration and scaling was done with the XDS and XSCALE software package (Kabsch, 2010 a). Initial phases were derived from single anomalous dispersion differences in diffraction data of crystals from SeMet-labeled protein or NH_4I-soaked crystals collected at the absorption edges of the corresponding elements. Heavy atom positions were identified with hkl2map (Pape et al., 2004) and forwarded to phenix.phaser of the PHENIX software suite (Adams et al., 2010). Initial model generation was performed with phenix.autobuild (Terwilliger et al., 2007) and followed by manual model building of poorly defined regions in Coot (Emsley et al., 2010) from the CCP4 software suite (Winn et al., 2011). Further refinement involved alternating rounds of manual adjustments and optimization in phenix.refine (Afonine et al., 2012).

3.7.2 NMR spectroscopy

NMR spectroscopy was performed with the help of Junior-Prof. Dr. Alexander Marchanka from the Institute of Organic Chemistry at the Leibniz University of Hannover. Characteristics of NMR spectroscopy were described in chapter 1.3.2. Simple 1D experiments provide information on the folding state of the protein. A folded, non-aggregated protein shows sharp peaks over a wide range of chemical shifts. The signals range from 0 ppm to 3 ppm for methyl groups and aliphatic residues, from 3.5 to 5.5 ppm for Hα's and up to 10 ppm for amine-associated, aromatic and backbone protons. Signals embedded in the ladder area give information on the tertiary structure and the folding state. A sample which provided a good 1D ^{1}H-spectrum was further addressed by ^{15}N-labelling and ^{15}N-^{1}H-HSQC, the so-called protein fingerprint (Cavanagh et al., 2007).

High protein purity was in general essential for interpretable spectra and was achieved by application of highly selective Strep-Tactin® purification and a SEC polishing step. For liquid state NMR, the protein needed to be concentrated to approximately 20 mg/mL and delivered in a small volume of 600 µL. The salt concentration was kept below 300 mM NaCl.

For an HSQC experiment, the protein was labelled with ^{15}N, according to the protocol described in chapter 3.5.4, and purified to homogeneity.

Samples for NMR were prepared freshly and freeze-thaw cycles were avoided. Samples were stored at 4°C and handed over promptly.

3.7.3 Small angle X-ray scattering

The sample purity and homogeneity are essential for measurements in batch mode, as the resulting signal will be a superimposition of all molecules dissolved. For this reason, the sample needed to be characterized in detail in terms of its stability and its behavior upon freeze and thaw cycles. A pure and homogeneous protein sample needed to be dialyzed o/n to enable background subtraction. Importantly, protein concentration after dialysis was avoided to circumvent possible aggregation. Instead, the samples were dialyzed at a high concentration and diluted to at least three working concentrations after dialysis. Typical working concentrations were between 20 mg/mL for small proteins, down to 1 mg/mL. A minimum sample volume of 50 µL was targeted. A buffer for background subtraction was provided in large excess.

For SEC-SAXS, separable impurities were tolerable, but reproducible elution profiles of a distinct species needed to be assured. The sample amount to load for the SEC run again correlated with the protein size. To maintain good resolution properties, 50 µL sample volume were not exceeded. The sample concentration was at least 5 – 7 mg/mL.

3.7.4 Electron microscopy

TEM was performed at the HZI with the help of Dr. Mathias Müsken from the Central Facility for Microscopy. Samples were delivered immediately after accomplishing the last purification step. Freeze-thaw cycles and high amounts of glycerol or salt were avoided.

Carbon coated copper grids with a mesh size of 400 (EMS, Hatfield, Pennsylvania, USA), as well as self-produced carbon foils were used for sample immobilization. A diluted droplet of 50 µL of the sample was placed on a clean, repulsive surface, such as parafilm, and two droplets of water and one droplet of a 4% uranyl acetate (UAc) solution were placed next to it. For adsorbance, the carbon foil or the carbon-coated grid was put on top of the sample droplet. Adsorbance time ranged from 15 seconds to 1 minute. Afterwards, the foil/grid was removed and excess liquid was carefully removed by touching the corner of the grid with filter paper. Subsequently the grid was washed twice by dipping it onto the water droplets. Staining was performed by placing the grid on the UAc droplet for 1 minute. The grid was then dried and ready for inspection.

Data collection was carried out at a LIBRA 120 electron microscope (Carl Zeiss AG, Oberkochen, Germany). Once an area suitable for data collection was found in the low magnification mode, the magnification was set to 20.000. Areas with a homogenous particle distribution were preferred. Images were taken, exported and converted to *.mrc* file format. Further processing was performed with the software package Relion.

4 Topic I: Architecture of the *Clostridioides difficile* proline reductase

Besides rudimentary knowledge of the molecular weight and partial identification of the active site, detailed understanding of the assembly and the architecture or structures of subunits of the proline reductase are lacking. This study aims to fill these gaps. A variety of techniques addressing diverse perspectives was applied and combined, including homologous and heterologous protein expression as well as biophysical characterization of isolated targets.

From the structural point of view, this task was split into a set of sub projects. Due to the assembly from rather small to medium-sized entities to the multimeric protein complex, versatile methods from the field of structural biology have been applied. The subunits PrdA and PrdB were addressed by X-ray crystallography. Due to its relatively small size of 23 kDa, the selenocysteine-containing subunit PrdB could also be subjected to NMR. Due to a size of around 870 kDa (Kabisch et al., 1999), the complete complex is a predestined target for TEM and cryoEM.

Heterologous production of recombinant PrdA and PrdB and affinity chromatography purification was performed to yield high amounts of unmatured protein. The interaction of both subunits was recovered and accurate size determination was performed by electrophoresis and scattering experiments. A similar set of techniques was applied to the natively purified, fully matured and completely assembled proline reductase complex after the purification protocol was modified to milder conditions.

The following chapters summarize the experiments that were applied to gather structural information of the proline reductase.

4.1 Proline reductase subunit PrdA

PrdA is produced as a proprotein with 67.6 kDa and is subsequently cleaved by a selenocysteino/cysteino/-serinolysis (dependent on the *C. difficile* strain) between amino acids 420 and 421, resulting in the two subunits PrdAβ with 44.6 kDa and PrdAα with 22.8 kDa. While the N-terminal subunit PrdAβ evolves as a regular C-terminus, the N-terminus of the C-terminal subunit PrdAα matures to a modified pyruvyl group, which is essential for proline binding. The supposed function of the subunits during the reaction cycle is shown in detail in chapter 1.1.4. Beside substrate binding and direct interaction with PrdB, further structure-function-relationships remain elusive. Here, the aim for detailed understanding of the proline reductase was set, which is why the gene *prdA* was cloned for recombinant protein production and characterization.

4.1.1 Cloning of prdA

Genes were amplified by PCR from gDNA of *C. diffcile* 630Δerm (van Eijk et al., 2015), provided by the DSMZ, Braunschweig. Correct restriction, insertion, ligation and transformation was reviewed by colony-PCR and sequencing. The detailed protocols are given in the chapters 3.4.3 to 3.4.7. As soon as correct cloning was verified, the purified DNA was transformed into *E. coli* expression strains and tested for expression.

4.1.2 Production of PrdA

Full length PrdA was produced in LB medium, following the protocol described in chapter 3.5.6, induced with 50 µM IPTG at 20°C. The protein was purified in a two-step purification procedure using Ni-NTA and SEC (S200 16/60). The preparative SEC showed strong protein aggregation as indicated by a main peak eluting in the exclusion volume at 45 mL. A second peak at 72 mL shows a much lower signal that might correlate to the size of a PrdA dimer. The SDS-PAGE showed exclusively a single band slightly above the 66 kDa marker protein. Hence proteolytic maturation towards PrdAα (22 kDa) and -β subunits (45 kDa) did not occur autocatalytically. Interestingly, most proteins that carry a pyruvoyl group act as decarboxylating or reducing enzymes on amino acids or amino acid derivatives, as the proline reductase does (Jiang et al., 2013). As described by Ute Kabisch and Brit Bednarski, the serinolysis of recombinant PrdA does not occur spontaneously in vitro but requires a large amount of strong reducing agent (Kabisch et al., 1999; Bednarski et al., 2001). Other examples are known for which the proteolytic maturation appears autocatalytically (Gallaghert et al., 1992; Albert et al., 1998; Tolbert et al., 2003; Bale et al., 2010). However, it is considered that in vivo, *C. difficile* provides chaperones for reduction and induction of serinolysis, as it has been shown for other proteins (Trip et al., 2011). Either way, it was suggested that a backbone oxygen, intrinsically present or provided by a chaperone, forms an oxyanion hole for stabilization of the intermediate oxyoxazolidine (Schmitzberger et al., 2003). Until today, only the *L*-aspartate-α-decarboxylase has been crystallized as a pyruvoyl-dependent protein in its inactive, unmatured state. From this study one can derive that the overall structural conformation is not affected by pyruvoyl formation. Yet, the direct environment shows reduced flexibility after maturation (Schmitzberger et al., 2003). Crystallization and other biophysical investigations of the purified full-length construct were not successful. The strong aggregation pointed towards misfolding of the protein. Nevertheless, the purified protein could be used to find truncation sites to break up the soluble agglomerates. Therefore, PrdA was subjected to limited proteolysis.

4.1.3 Limited proteolysis of PrdA

The soluble but aggregated protein was treated by limited proteolysis in order to determine boundaries for soluble PrdA-constructs. The aim was to find a soluble PrdA construct forming a defined oligomeric state. Hence PrdA was incubated with various protease and PrdA fragments from different time points were separated on SDS-PAGE. While no digestion with chymotrypsin, papain, endoproteinase glu-C, clostripain, bromelain and actinase E was observed, trypsin, subtilisin and proteinase K degraded the protein completely without generating any distinct band for further analysis.

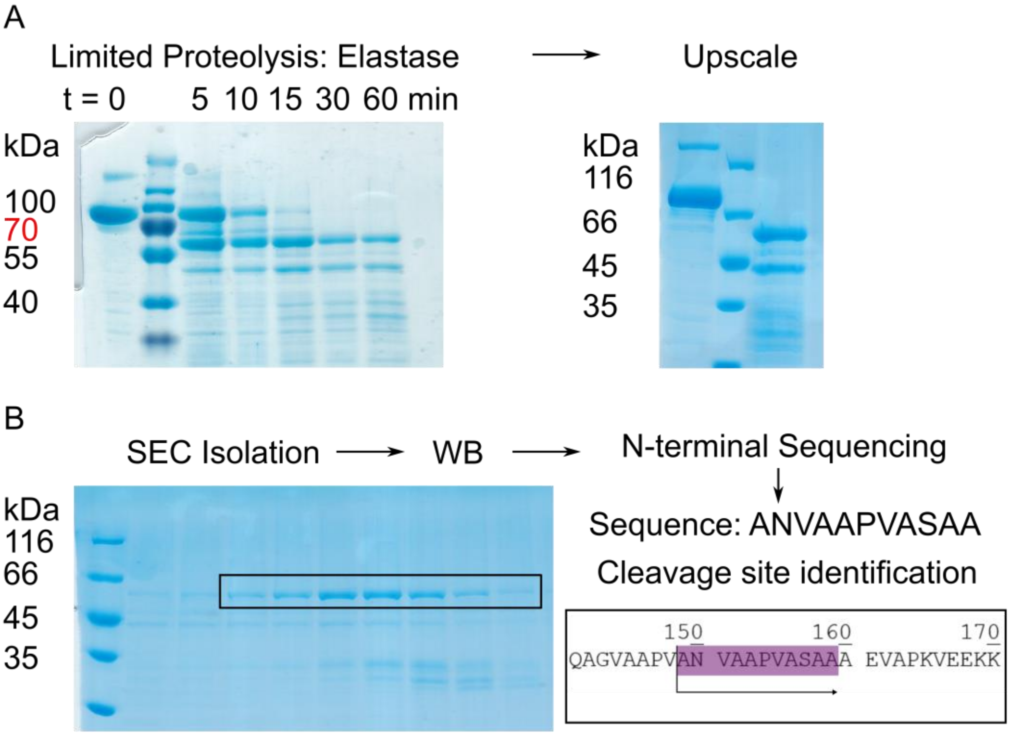


Figure 4.1: Limited proteolysis of PrdA. A) PrdA formed soluble aggregates and could not be used for bioanalytic analysis. Therefore, the protein solution was mixed with elastase and digested at 37°C. SDS-PAGE samples were taken after 5, 10, 15, 30 and 60 minutes. As the protein band corresponding to PrdA was digested and formed bands corresponding to lower molecular weight, the reaction was upscaled. B) The reaction product was applied to an analytical SEC (S200 10/30 increase) and a soluble species was isolated. The sample was transferred on a PVDF membrane by western blot and N-terminal sequencing identified the cleavage site A149 of the N-terminus.

Other proteases like trypsin, subtilisin and proteinase K completely degraded the protein without generating distinct bands for further analysis. Pepsin, cleaving N-terminally of phenylalanines, shows a gradual proteolysis with many bands appearing. Due to difficult separation of the resulting bands and the mostly internally located phenylalanines, Pepsin was dismissed from the selection. However, limited proteolysis by elastase yielded two distinct bands with lower molecular weight (see figure 4.1). To determine the exact cleavage site of elastase, which is predicted to cleave adjacent to neutral amino acids, an upscaled batch of PrdA-elastase mix was prepared and subsequently analyzed by analytical SEC (S200 10/30) as depicted in figure 4.1. The composition for the single fractions collected was visualized by SDS-PAGE. Fraction 25, eluting at 13 mL, contains exclusively one digestion product with a slightly reduced MW compared to full length PrdA. This material was selected for N-terminal sequencing. The sample was applied to SDS-PAGE and transferred onto a PVDF-membrane by WB. The transferred band was cut and analyzed by N-terminal sequencing. N-terminal sequencing clearly located the elastase cleavage site at position A149. According to bioinformatical predictions (PSIPRED, DISOPRED3) the elastase cleavage site is located in one of two stretches that are presumably highly flexible (compare figure 4.2).

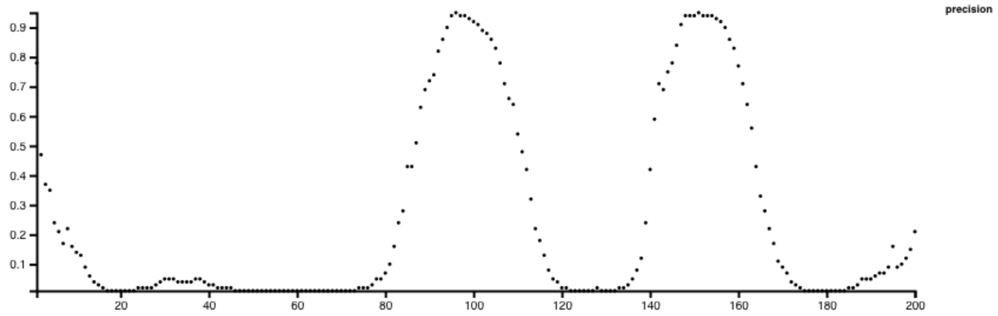


Figure 4.2: Disorder prediction for the N-terminal sequence stretch of PrdA. In the range from amino acid 80 to 120 and 140 to 170, the protein is predicted to be highly flexible by PSIPRED DISOPRED3. The elastase cleavage site is located in the center of the second flexible region at A179.

It has been described before that protein interaction sites are often formed by flexible regions that adopt a more rigid and stable conformation upon binding (Dunker et al., 2005). Hence, it appears likely that the flexible region 1 – 149 potentially represents an oligomerization site. As the cleavage site at A149 was found to lead to a soluble construct of PrdA with a defined oligomerization state, the truncated construct PrdA$_{149-626}$ was cloned and expressed.

4.1.4 Production of N-terminally truncated PrdA

In the following, N-terminally truncated versions of PrdA were cloned, expressed and purified by a two-step purification protocol using Ni-NTA and SEC (S200 26/600). PrdA$_{149-626}$ is the soluble construct that was identified by limited proteolysis. PrdA$_{158-626}$ starts at the predicted helix6 and was additionally designed from the sequence-based SS prediction. Both constructs yielded soluble protein. This leads to the assumption that the 148 mostly flexible N-terminal residues cause the observed heavy agglomeration. In direct comparison, PrdA$_{158-626}$ shows a non-symmetrical elution profile in SEC, whereas PrdA$_{149-626}$ elutes in one predominant symmetrical peak. Therefore, PrdA$_{149-626}$ was used for further analysis.

4.1.5 Thermal shift assay with PrdA

A set of thermal shift assays for PrdA$_{149-626}$ were performed to determine optimal buffer conditions for purification, storage and crystallization. The protein was mixed with SYPROOrange and the additives as described in chapter 3.6.1. As one can derive from figure 4.3, the protein boundaries identified from limited proteolysis and N-terminal sequencing (see chapter 3.5.13) led to a construct that is soluble, stable and therefore suitable for bioanalytical techniques. In the storage buffer PrdA$_{149-626}$ shows a melting temperature of 51°C. The first derivative of the melting curve (see figure 4.3) possesses a symmetric minimum, which indicates that this construct forms one entity and partial denaturation of several subunits does not occur.

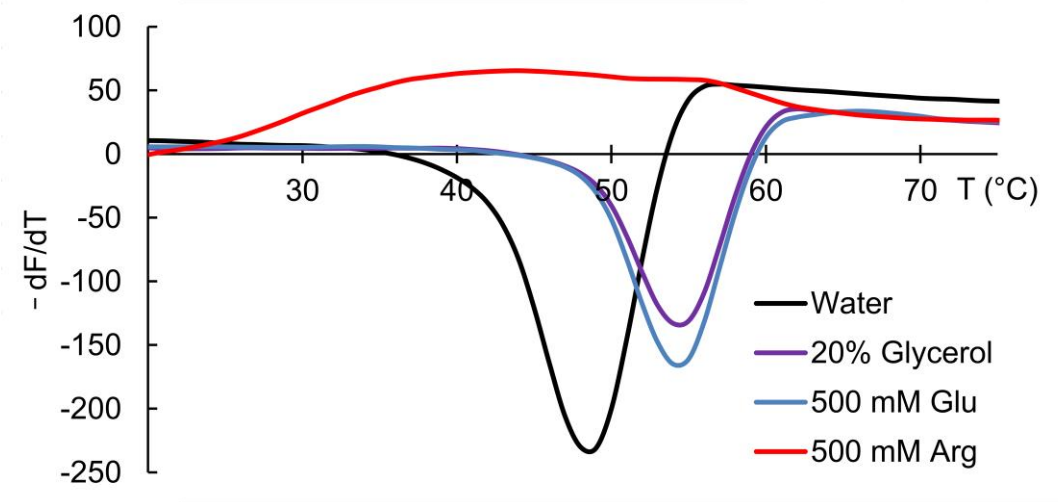


Figure 4.3: Melting profiles for PrdA$_{149-626}$. Diluting the sample with water serves as a reference. In direct comparison, 20% glycerol and 500 mM glutamate significantly stabilize the protein by increasing the melting temperature from 51°C to 56°C. On the other side, addition of 500 mM arginine led to complete diminishing of the fluorescence signal. Melting temperatures are given in table 4.1.

Addition of 20% glycerol showed a strong positive effect, which was exhibited by an increased melting temperature of 56°C. The same observation was made for the addition of 500 mM glutamate. Both chemicals can serve as protein stabilizers by shielding hydrophobic surface patches. Ionic interactions to positively charged surface residues are likely. However, upon addition of 500 mM arginine, a positively charged and chaotropic amino acid, PrdA$_{149\text{-}626}$ does not show any melting signal, which indicates complete denaturation of the sample. Melting temperatures are given in table 4.1.

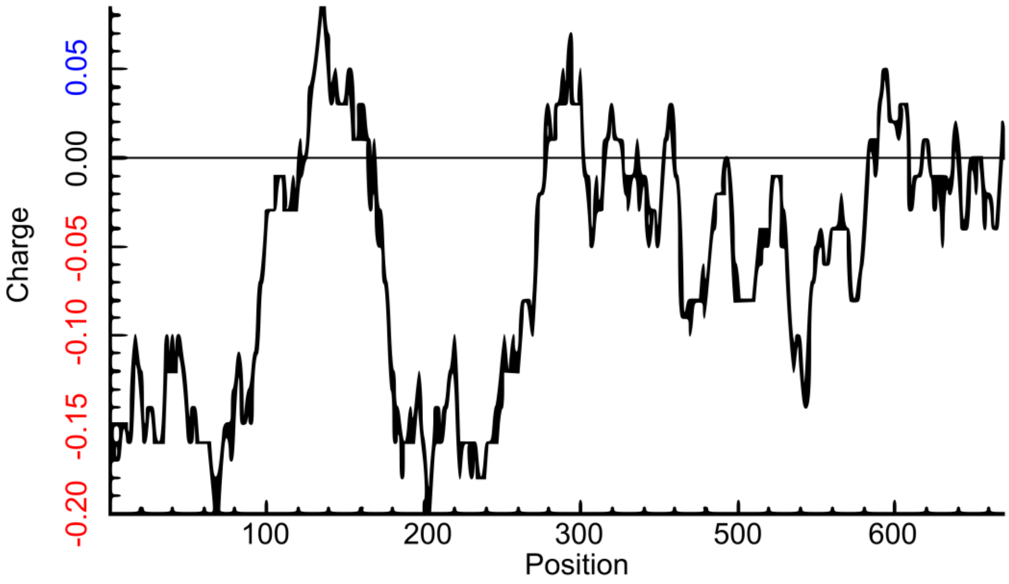


Figure 4.4: EMBOSScharge analysis of PrdA. The sequence of PrdA was analyzed in terms of net charge over amino acid position. This plot shows abnormally low values for the sequence stretches 1 - 100 and 180 – 280. The local minima are framing a region of strong positive net charge for the amino acid positions 100 – 180.

Considering that the net charge difference of glutamate and arginine might have played a major role in the observed difference in results, the sequence of PrdA was analyzed in terms of its charge over amino acid position. This analysis showed obvious commonalities between sequence-based flexibility analysis and net charge. The first stretch in the PrdA sequence shows a low net charge, as one can derive from figure 4.4. The termination of this glutamate rich region is accompanied by a high net charge in the stretch from amino acid position 100 to 180. In this region the elastase cleavage site A149 is located. This could signify that the removed domain is comprising an oligomerization site. Interestingly, the first stretch of negative net charge is followed by a second stretch that ranges from amino acid position 180 – 280. Observing the loss of structural order and considering that this second negatively charged stretch forms another interaction site, it can be hypothesized that this region represents another protein-protein interface. An interaction between proteins with glutamate- and aspartate-rich regions has already been reported for the hirudin-thrombin-interaction (Sharp, 1996).

In this case, addition of high concentrations of arginine could potentially disrupt ionic interactions in an interface and lead to denaturation of PrdA$_{149-626}$. An important information for this hypothesis is the definition of the oligomerization state of this construct.

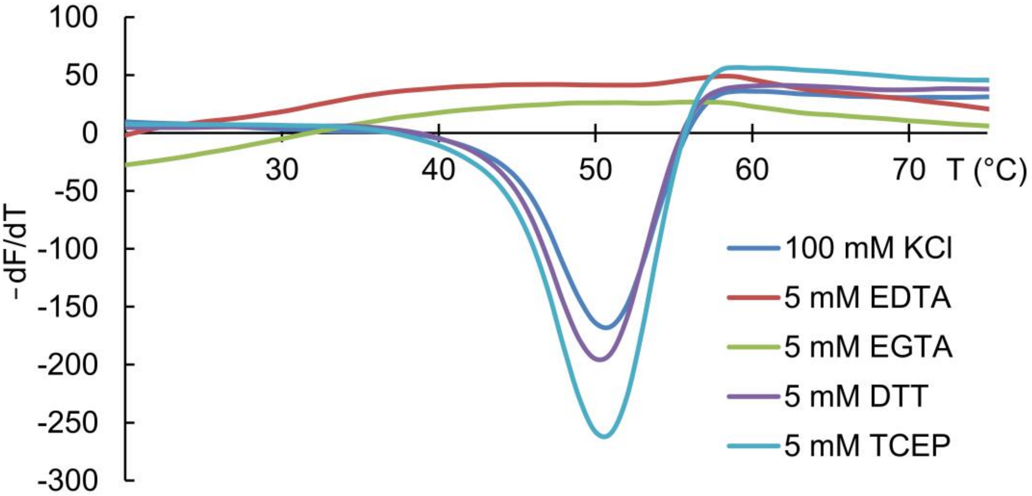


Figure 4.5: Melting profiles for PrdA$_{149-626}$. Comparison of melting curves after addition of KCl, chelators and reducing agents. PrdA149-626 does not show any significant stabilization or destabilization upon addition of KCl, DTT or TCEP. However, adding chelators such as EDTA and EGTA, the fluorescence signal is completely diminished. Melting temperatures are given in table 4.1.

Further information can be extracted from figure 4.5, which shows that reducing agents such as DTT and TCEP do not affect protein stability, despite the fact that it has previously been shown that DTT can be used to induce proline reductase activity and serve as an electron donor (Kabisch et al., 1999). However, addition of chelating agents, such as EDTA and EGTA, led to complete extinction of the melting signal. This finding indicates that ions may play a critical role in protein stability. As soon as those ions were removed from PrdA$_{149-626}$, the protein denatured. Calculated melting temperatures T_m are given in table 4.1.

Table 4.1: Melting temperature for $PrdA_{149\text{-}626}$. Melting temperatures were determined in a thermal shift assay applying the RUBIC Additive Screen (Hampton Research, Aliso Viejo, California, USA). Corresponding melting peaks are shown in figures 4.3 and 4.5.

Additive	Tm [°C]	Additive	Tm [°C]
Control (water)	51	1 mM $ZnCl_2$	42
20 % glycerol	54	5 mM EDTA	not detectable
500 mM Glutamate	56	5 mM EGTA	not detectable
500 mM Arginine	not detectable	5 mM DTT	50
100 mM KCl	51	5 mM TCEP	51

The presence of metal ions has been described to be essential for protein folding in many cases before. For instance, it was found that the intrinsically disordered protein Prothymosin-α requires zinc to establish protein-protein interactions. Interestingly, Prothymosin-α exhibits a glutamate-rich region, just as PrdA does in regions with significant low net charge (Garapati et al., 2018). It has also been described that zinc and other metal ions were bound by acidic binding motifs (Herrera et al., 2019). Indeed, it was observed that zinc plays a role in proline reductase melting behavior. However, in terms of addition of zinc, TSA results were inconsistent and varied from a positive effect of stabilization of the proline reductase supercomplex by ΔTm of 13°C (data not shown) to a destabilization by 9°C for $PrdA_{149\text{-}626}$ (compare table 4.1). This suggests that the interaction of proline reductase with zinc might be multi- or cross-functional and thereby hard to monitor.

4.2 Proline reductase subunit PrdB

PrdB is the 25.7 kDa subunit of the proline reductase complex that contains the active-site selenocystein-cystein, predicted to induce the reductive ring cleavage of *D*-proline. For PrdB, neither the structure nor protein interfaces to interaction partners are predictable.

4.2.1 Cloning of prdB$_{U151C}$

The sequence of PrdB was amplified from the gDNA of *C. difficile* 630Δ*erm* (van Eijk et al., 2015). As PrdB has a seleno-cystein at amino acid position 151, which is encoded by an alternative stop codon in *E. coli* expression strains, this position was mutated into a cysteine using PIPE cloning ($prdB_{U151C}$). The mutated gene $prdB_{U151C}$ was subsequently cloned into different expression vectors. Success of cloning of the PrdB$_{U151C}$ mutant was monitored by agarose gel electrophoresis and DNA sequencing as described in chapters 3.4.2 to 3.4.7.

PrdB could be produced with both TEV cleavable N-terminal Strep-tag and C-terminal His$_6$-tag. Either way, a large amount of pure PrdB could be isolated. The elution volume from a SEC (S75 16/60) suggests PrdB to be present as a monomer.

4.2.2 Production of PrdB$_{U151C}$

PrdB could be expressed with both, a TEV cleavable N-terminal Strep-tag and C-terminal His$_6$-tag. Either way, a large amount of pure PrdB could be yielded. The elution volume from a SEC (S75 16/60) suggests PrdB to be present as a monomer.

4.2.3 Thermal shift assay with PrdB$_{U151C}$

For buffer optimization, PrdB was diluted to 1.5 mg/mL and mixed with SyproOrange and the stock solutions as described in chapter 3.6.2. In a first attempt, the melting temperature was determined for a variety of buffers, covering a range from 0 to 500 mM NaCl and a pH range from 4.0 to 10.5.

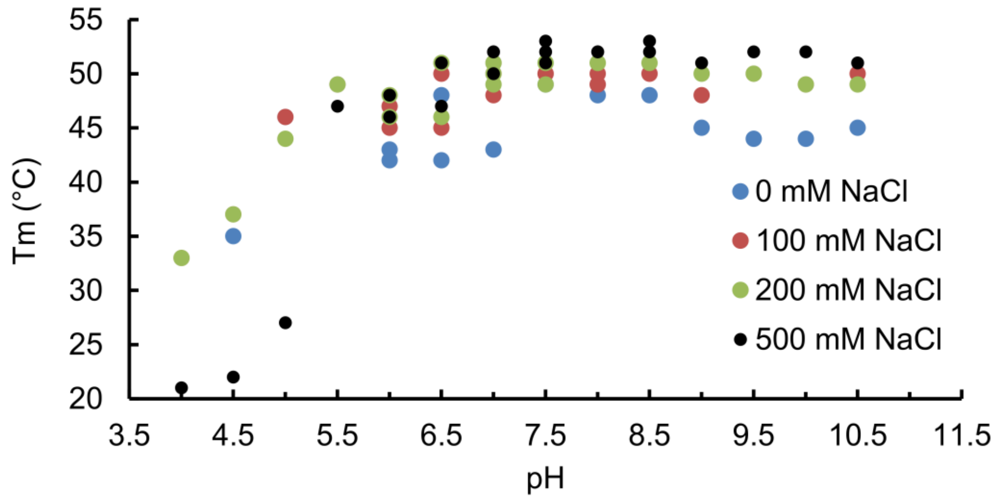


Figure 4.6: Thermal Shift Assay PrdB$_{U151C}$. The melting temperature for the protein is shown in dependency of the salt concentration (blue, red, green, black) and the pH tested. In general, highest values for Tm were achieved with a pH between 7 and 8.5. Increasing the salt concentration, gradually led to an increased Tm.

As can be derived from figure 4.6, the optimal pH for PrdB$_{U151C}$ ranges from 7.0 to 8.5. In this area the protein tolerates both, high and low salt concentrations. Interestingly, if the pH value drops below the theoretical pI of 5.7, there is a significant drop of Tm when high concentrations of NaCl are present. The effect is less pronounced for lower salt concentrations. The maximum melting temperature was 54°C.

Subsequently, an additive screening was performed to identify other stabilizing compounds. Unsurprisingly, this experiment showed that glycerol stabilizes the protein, but no other stabilizing buffer components could be identified.

4.2.4 NMR with PrdB$_{U151C}$

With a MW of 25.7 kDa, PrdB$_{U151C}$ is a suitable candidate to be addressed by NMR for structural determination. Collection and interpretation of the spectra was performed by Prof. Dr. Alexander Marchanka from the LU Hannover.

For these measurements, the protein was expressed in a 2 L TB fermenter batch, which yielded approximately 40 mg of pure PrdB$_{U151C}$. The protein was concentrated to 20 mg/mL and dialyzed against 100 mM NaCl, 20 mM HEPES pH 8.5. The spectrum shown in figure 4.7 is characterized by sharp signals for N-H coupling with a chemical shift of approximately 6.5 – 9.5 ppm. The C$_\alpha$-H coupling is shown with a shift of 3 - 5 ppm, while side chain protons rise in the region of 0 – 5 ppm.

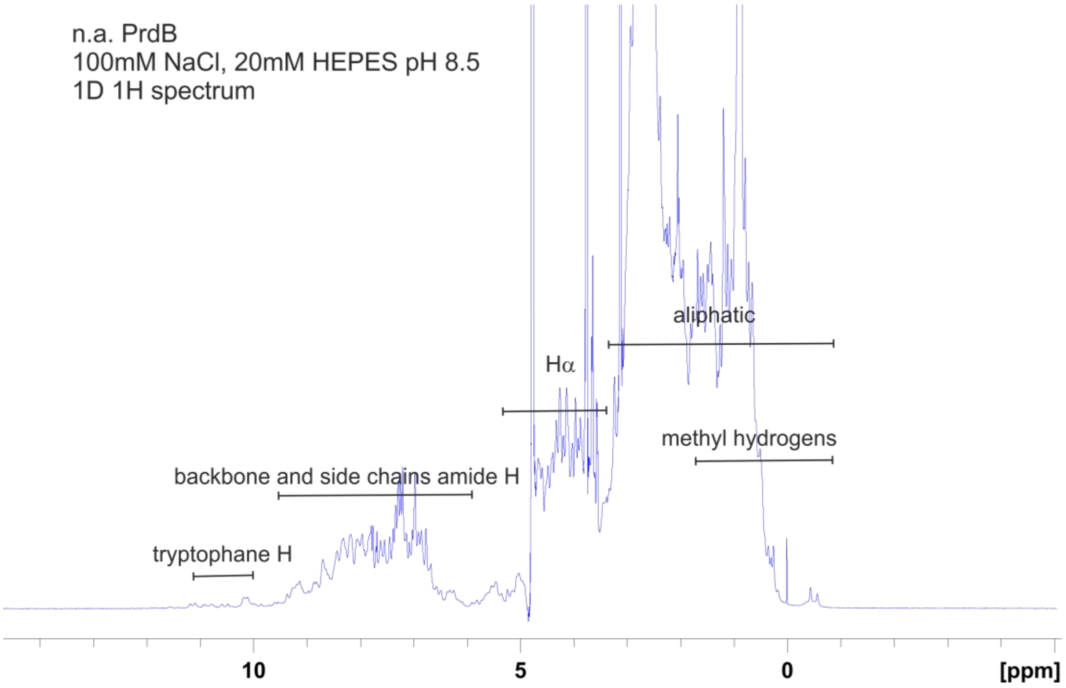


Figure 4.7: 1D NMR spectrum of $\text{PrdB}_{\text{U151C}}$. This first experiment shows sharp signal lines for characteristic couplings in proteins. Typical ranges for chemical shifts that occur frequently in proteins are marked by black bars.

This first NMR experiment shows that $\text{PrdB}_{\text{U151C}}$ is a monomer and did not aggregate during the measurement, as the peaks are well resolved. This is a basic requirement for successful structure determination by NMR.

Hence, $\text{PrdB}_{\text{U151C}}$ was labelled with ^{15}N during another fermenter batch in M9MM. By replacing the natural abundant ^{14}N with ^{15}N, the signal for correlations is sharpened, as quadrupolar interactions that appear for ^{14}N are erased. This way N-H correlations become interpretable. In a 2D HSQC, a folded, monodisperse protein typically shows a signal for each backbone N-H and additional signals for each N-H group present in sidechains. As asparagines and glutamines hold a primary amine group, each sidechain raises two peaks with the same shift for the ^{15}N atom (110 - 115 ppm) and two shifts for the two protons (6 - 8 ppm). $\text{PrdB}_{\text{U151C}}$ possesses 9 Asn and 7 Gln, potentially leading to 32 additional signals in the mentioned region of the spectrum. The ^{15}N that are incorporated into tryptophane's indole ring resides in a chemical environment that differs from that of amines. Therefore, their chemical shifts differ strongly. Usually, signals for Trp can be found in a range of 125 - 125 ppm for ^{15}N and 10 - 11 ppm for protons. As the PrdB sequence has only one Trp, there should be only one additional signal. With 241 backbone N-H pairs and described backbone couplings, $\text{PrdB}_{\text{U151C}}$ is expected to show 274 peaks in an HSQC protein fingerprint. However, this experiment showed more signals than expected, with a plurality of weak signals scattered over a wide range of chemical shifts, attesting some structural inconsistencies, as one can see in figure 4.8.

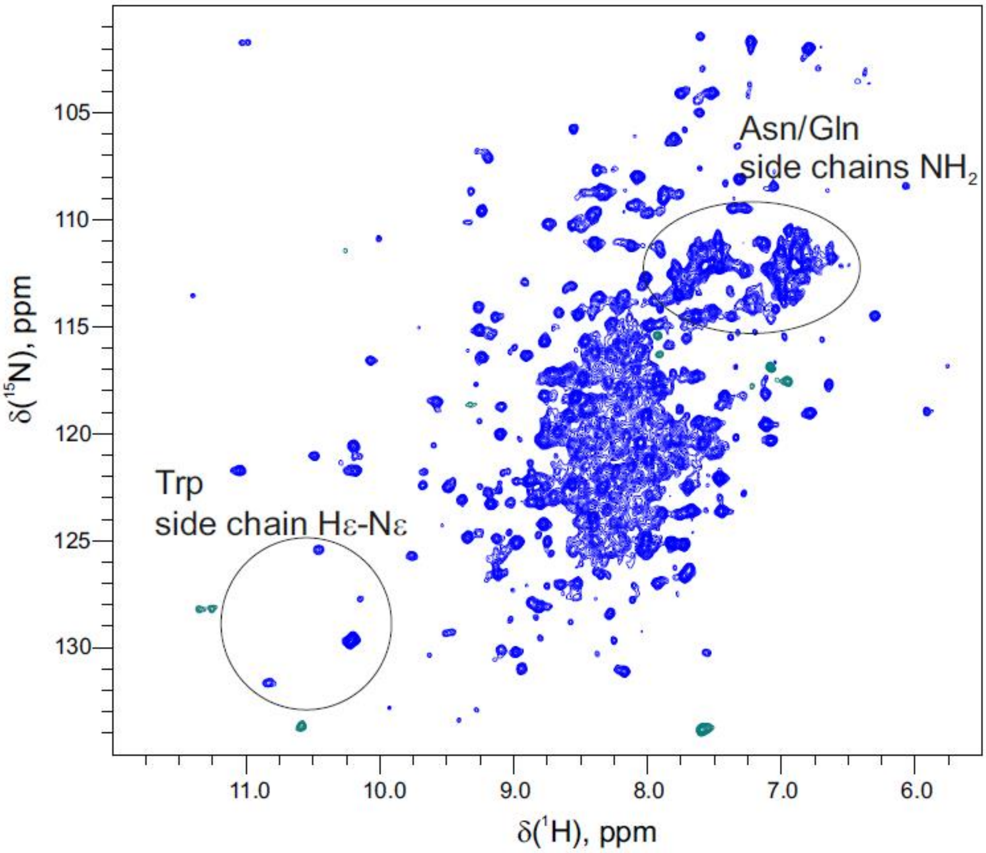


Figure 4.8: 2D NMR spectrum of $PrdB_{U151C}$. The HSQC or protein fingerprint. The sample shows a well-defined behavior in terms of solubility and monodispersity, as the obtained peaks show sharp signals. However, the number of signals detected exceeds expectations.

As the protein was tested at high purity, it can be excluded that the additional signals originate from impurities. Further, dimerization appears unlikely, as signals remain as sharp as before. This would not be the case for an oligomerizing or aggregating sample, which would show much broader peaks due to higher rotational correlation time, which enhances relaxation speed. Therefore, it was concluded that recombinant PrdB might possess different conformations. This assumption was investigated using CD spectroscopy and fluorescence-based melting temperature determination, both techniques capable to distinguish between different protein species.

4.2.5 CD spectroscopy with PrdB$_{U151C}$

Previous NMR experiments suggested that recombinant PrdB$_{U151C}$ might show several conformations. As this circumstance likely affects the SS content, CD spectroscopy was chosen to find further evidence. During SEC of PrdB$_{U151C}$-His$_6$, the main peak corresponding to PrdB$_{U151C}$ was flanked by a shoulder, indicating the presence of two species with variant radii of gyration. Careful fractionation of eluting species made comparable studies on early and late eluting material possible, although both species could not be separated completely. Collected samples were concentrated to 1 mg/mL and dialyzed against PrdB storage buffer before subjection to CD spectroscopy. Results are shown in figure 4.9.

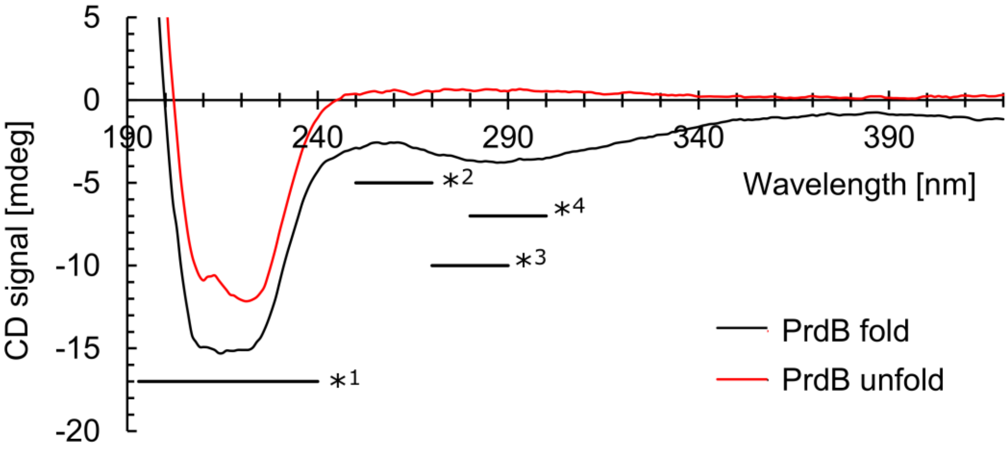


Figure 4.9: CD spectroscopy with PrdB$_{U151C}$. The direct comparison of two spectra on PrdB$_{U151C}$, corresponding to an early and late eluting species shows clear differences. The CD spectrum can be divided in several areas or interest, *1 marks the range from 190 to 240 nm, which gives information on the SS of the sample. From 250 to 270 nm, phenylalanines contribute to the signal, indicated by *2. In the following range, *3 marks 270 to 290 nm, where tyrosine signals can be found and finally *4 ranging from 280 to 300 nm describes tryptophane signals in a chiral environment. Folded PrdB$_{U151C}$ shows a clear signal with -4 mdeg at 290 nm. This signal cannot be found for the other PrdB$_{U151C}$ species.

Comparison of the spectra for both species shows significant differences in secondary structure composition. SS elements give characteristic minima at 208 and 220 nm for α-helices and at 215 nm for β-strands, which are differently pronounced. Additionally, comparison of the near UV range (250-300 nm) shows a change for signals induced by aromatic residues that are embedded in a chiral environment. The near UV CD signal is dependent on protein folding, rigidity and amino acid composition. PrdB$_{U151C}$ sequence analysis suggests that the signal difference at 290 nm arises from a changed environment for the only tryptophane that is present in PrdB$_{U151C}$. As the signal giving tryptophane is located at amino acid position 25, following immediately on a triple-proline-motif, an increased flexibility of the N-terminus could cause that observation.

4.2.6 Intrinsic fluorescence thermal shift assay with PrdB$_{U151C}$

Previous experiments showed a difference for the CD signal of aromatic residues (Chapter 4.2.5) as well as the presence of multiple conformations in NMR studies of PrdB$_{U151C}$ (Chapter 4.2.4). In the following intrinsic fluorescence thermal shift assay (intrinsic fluorescence TSA), a tryptophane fluorescence shift (350 nm → 330 nm) is caused by a change in the chemical environment of the sidechain. Typically, the change is part of an unfolding event and can therefore be used to calculate the Tm (Moon et al., 2011). Again, the two species of PrdB$_{U151C}$, separated by SEC, were compared. The plotted ratios show a clear difference, as one can see in figure 4.10.

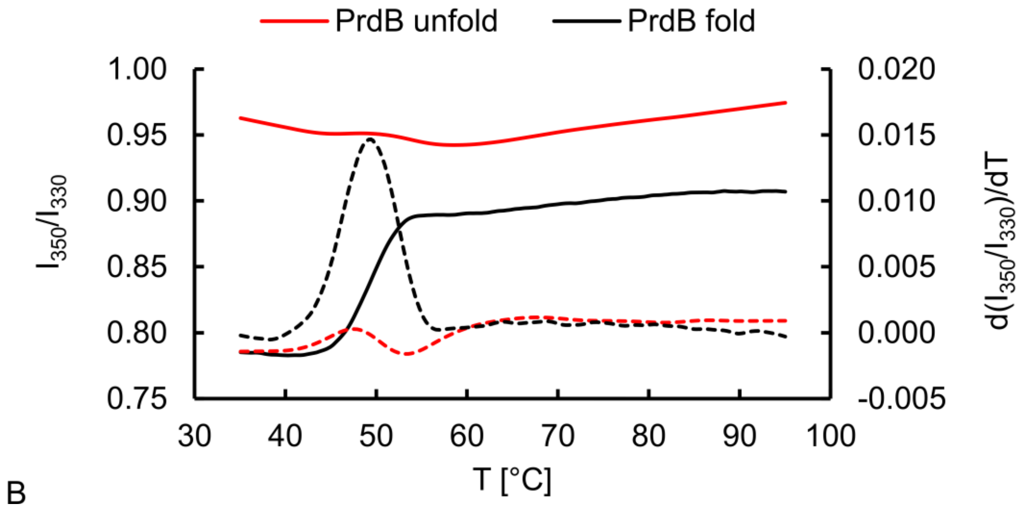


Figure 4.10: Intrinsic fluorescence TSA with PrdB$_{U151C}$. Two separated species of PrdB$_{U151C}$ are directly compared by melting the proteins and monitoring the ratio of fluorescence signals at 350 and 330 nm. For the folded PrdB$_{U151C}$ (black), a typical curve can be observed were the ratio increases upon denaturation. The unfolded PrdB$_{U151C}$ (red) shows an increased ratio from the very beginning of the experiment. The 1st derivatives of the melting profiles are represented as dashed lines (black and red).

A high 350 nm/300 nm-ratio at the beginning of the experiment (red line) indicates that the tryptophane sidechain is already surrounded by a hydrophilic environment, whereas the curve of the second species (black line) shows a transition for the tryptophane residue, indicating translocation of the sidechain from a hydrophobic core to the hydrophilic solvent upon heat induced denaturation.

As the signal is based on Trp emission, this experiment confirms, that the Trp is directly affected by the conformational change of PrdB$_{U151C}$. A closer look at the primary structure of PrdB$_{U151C}$ shows a triple-proline motif, followed by the only tryptophane present in PrdB$_{U151C}$ at position 25. This peculiar composition is likely to form disordered structures as predicted by sequence based structural disorder estimations (DISOPRED). So far, it is not clear whether both conformations are capable to interact with PrdA.

Figure 4.11: N-terminal consensus sequence of PrdB. The sequence ranging from Val17 to Leu32 shows a conspicuous high number of prolines (red), presumably causing a high flexibility in this region. Furthermore, a tryptophane (green) is embedded in this area. It is remarkable that this residue represents the only tryptophane in the whole PrdB sequence.

At present, the uncontrollability of conformational heterogeneity impedes biochemical and biophysical characterization of PrdB$_{U151C}$. Therefore, understanding the cause for this heterogeneity is of critical importance. Clearly, structural elucidation of PrdB$_{U151C}$ by NMR is exclusively achievable after elimination of that behavior. It is likely that a mixture of conformational states prohibits crystallization of PrdB$_{U151C}$ and is the cause for poor reproducibility in PrdA-PrdB-interaction studies. Nevertheless, the intrinsic fluorescence TSA provides an opportunity to screen a large number of purified protein batches within a short time, making this goal achievable.

4.3 Interaction of recombinant PrdA$_{149-626}$ and PrdB$_{U151C}$

The isolation of a suitable fragment of PrdA (PrdA$_{149-626}$, chapter 4.1.3) enabled experiments to further investigate the interaction between PrdB and PrdA. Both proteins, PrdA$_{149-626}$ and PrdB$_{U151C}$ were expressed and purified separately, mixed in an equimolar ratio and subsequently characterized.

4.3.1 SEC-MALS with PrdA$_{149-626}$/PrdB$_{U151C}$

PrdA$_{149-626}$ and PrdB$_{U151C}$ were characterized by SEC-MALS and tested for interaction. For each run, 8 nmol of protein sample were loaded. The injection volume was 44 µL. The mixture of PrdA$_{149-626}$/PrdB$_{U151C}$ was set up using 34 µl of PrdA$_{149-626}$ and 10 µl of PrdB$_{U151C}$.

Table 4.2: SEC-MALS experiments for PrdA and PrdB. Experiments were set up by mixing of PrdA$_{149-626}$ and PrdB$_{U151C}$ in an equimolar ratio. The total injection volume was kept constant with 44 µL. The final concentration of injected sample was targeted to be 182 µM, resulting in an amount of 8 nM of protein for each run.

	MW (kDa)	c (mg/mL)	c (µmol/l)	$V_{PrdA\ 149-626}$ (µl)	$V_{PrdB\ U151C}$ (µl)	V_{buffer} (µl)	Inj. mass (µg)	c_{final} (µM)
PrdA$_{149-626}$	52.5	12.4	235.6	34	0	10	420.6	182.1
PrdB$_{U151C}$	25.7	20.6	799.2	0	10	34	205.5	181.6
PrdA$_{149-626}$B$_{U151C}$	78.2	14.2	182.0	34	10	0	626.1	182.0

The SEC-MALS experiment was performed with a S200 10/30 *increase* column at a flow rate of 0.5 mL/min. Each injection had a volume of 44 µL. Data collection and analysis was performed using the ASTRA® software package, version Astra 7.3.2.19 (Wyatt Technology Corporation, USA). Determination of protein concentration is based on RI-measurement. The refractive index increment dn/dc was considered to be 0.185 mL/g for proteins.

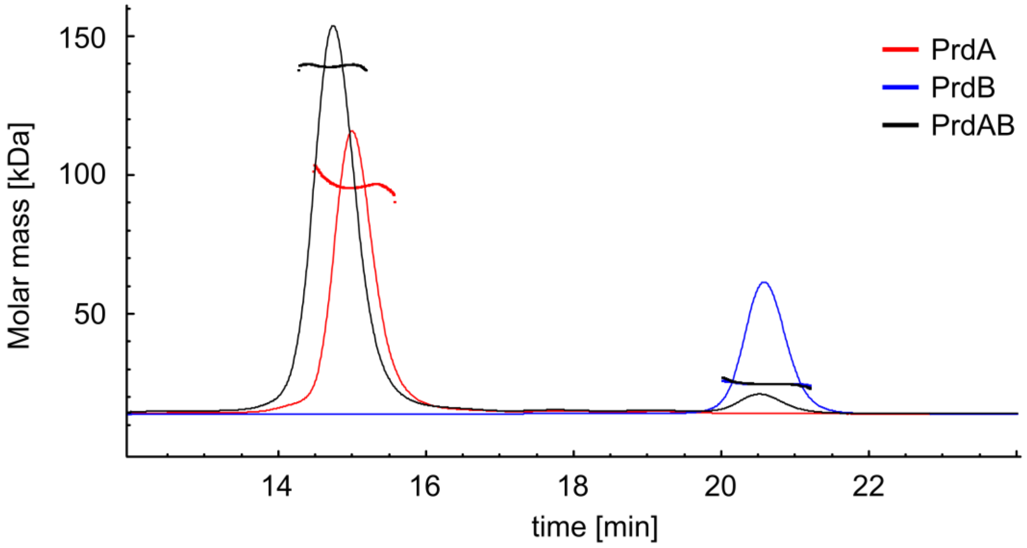


Figure 4.12: SEC-MALS measurement of PrdA and PrdB. The chromatogram shows the refractive index (RI) of the complex PrdA149-626/PrdBU151C (black line), PrdA149-626 (red line) and PrdBU151C (blue line). The calculated molar masses of the different species are shown as dots (g/mol). The eluting species PrdA and PrdB are well separated and show a distinct elution volume. The mixed species PrdAB shows a shifted elution volume (14.7 min), compared to PrdA, indicating a protein-protein interaction.

The chromatogram shows well-defined elution profiles for all species. After mixing $PrdA_{149\text{-}626}$ and $PrdB_{U151C}$, the elution is shifted towards higher MW, indicating a protein-protein interaction. The MW is calculated from the measured RI. Results are summarized in table 4.3.

Table 4.3: SEC-MALS results for PrdA and PrdB. The interaction of $PrdA_{149\text{-}626}$ and $PrdB_{U151C}$ was investigated by SEC-MALS. The recombinant samples were analysed as separately purified proteins and *in-vitro* mixed batches.

Sample	theo. MW monomeric (kDa)	Inj. mass (µg)	Mass recovery (%)	calc. MW (kDa)	Uncertainty (%)	Oligo factor
PrdA	52.5	420.6	73.1	96.3	0.6	1.8
PrdB	25.7	205.5	74.5	24.9	0.6	1.0
PrdAB	78.2	626.1	71.2	139.4	0.6	1.8

The calculated MW for $PrdA_{149\text{-}626}$ is 96.3 kDa. With an expected MW for 52.5 kDa for a monomer, this result suggests $PrdA_{149\text{-}626}$ is present as a dimer. $PrdB_{U151C}$ on the other hand elutes after 20.6 min and has a calculated MW of 24.9 kDa, which is consistent with a monomer with an expected size of 25.7 kDa. Mixing both proteins in equimolar amounts leads to a shift of the elution volume of the sample to 14.7 min. Additionally, a minor peak for free $\mathrm{PrdB}_{\mathrm{U151C}}$ was observed. The main peak was determined to have a MW of 139.4 kDa. The theoretical masses for the heterotetramer $(\mathrm{PrdA}_{149\text{-}626})_2/(\mathrm{PrdB}_{\mathrm{U151C}})_2$ are 156.4 kDa, while a heterotrimer $(\mathrm{PrdA}_{149\text{-}626})_2/\mathrm{PrdB}_{\mathrm{U151C}}$ adds up to 130.7 kDa. Although the determined mass of 139.4 kDa is closer to a heterotrimer, the complex is thought to be present as a heterotetramer, as the amount of free $\mathrm{PrdB}_{\mathrm{U151C}}$ was decreased to an extend that rules out a 2:1 complex. Further, the determined mass for the $\mathrm{PrdA}_{149\text{-}626}$ dimer is already settled 9% below the expected mass, which can be due to further terminal degradation or measuring inaccuracy. Assuming 96.3 kDa to be the true value for the MW of the $\mathrm{PrdA}_{149\text{-}626}$ dimer, a heterotetramer would result in a complex of 147.7 kDa, which is closer to the results as a heterotrimer. Also, gel band intensities on SDS-PAGES suggest a 2:2-ratio. Therefore, it can be concluded that $\mathrm{PrdA}_{149\text{-}626}$ and $\mathrm{PrdB}_{\mathrm{U151C}}$ form a hetero tetramer.

4.3.2 SEC-SAXS with $\mathrm{PrdA}_{149\text{-}626}/\mathrm{PrdB}_{\mathrm{U151C}}$

Further, a SEC-SAXS experiment with a multi-detector setup was performed. The synchrotron SAXS data were collected at beamline P12 operated by EMBL Hamburg at the PETRA III storage ring at the DESY, Hamburg (Blanchet et al., 2015). The purified samples were shipped to the synchrotron in a shock-frozen state on dry ice. They were stored at -80°C and thawed when needed.

To enhance complex formation $\mathrm{PrdA}_{149\text{-}626}/\mathrm{PrdB}_{\mathrm{U151C}}$ and $\mathrm{PrdB}_{\mathrm{U151C}}$ were crosslinked using 15 mM DSSO. $\mathrm{PrdA}_{149\text{-}626}$ was not crosslinked, as the dimer reacted sensitively to the reagent with complex decomposition. The scattering was used to determine I_0, R_g, the MW, the Porod volume and D_{max}, as well as to generate an ab initio bead model of the proteins. Native $\mathrm{PrdA}_{149\text{-}626}$ was applied with 1.237 mg. $\mathrm{PrdB}_{\mathrm{U151C}}$ was crosslinked with 15 mM DSSO and applied with 1.661 mg. $\mathrm{PrdA}_{149\text{-}626}\mathrm{PrdB}_{\mathrm{U151C}}$ was also crosslinked and applied with 1.64 mg. Scattering elution profiles and subtracted raw scattering is shown in figure 4.13.

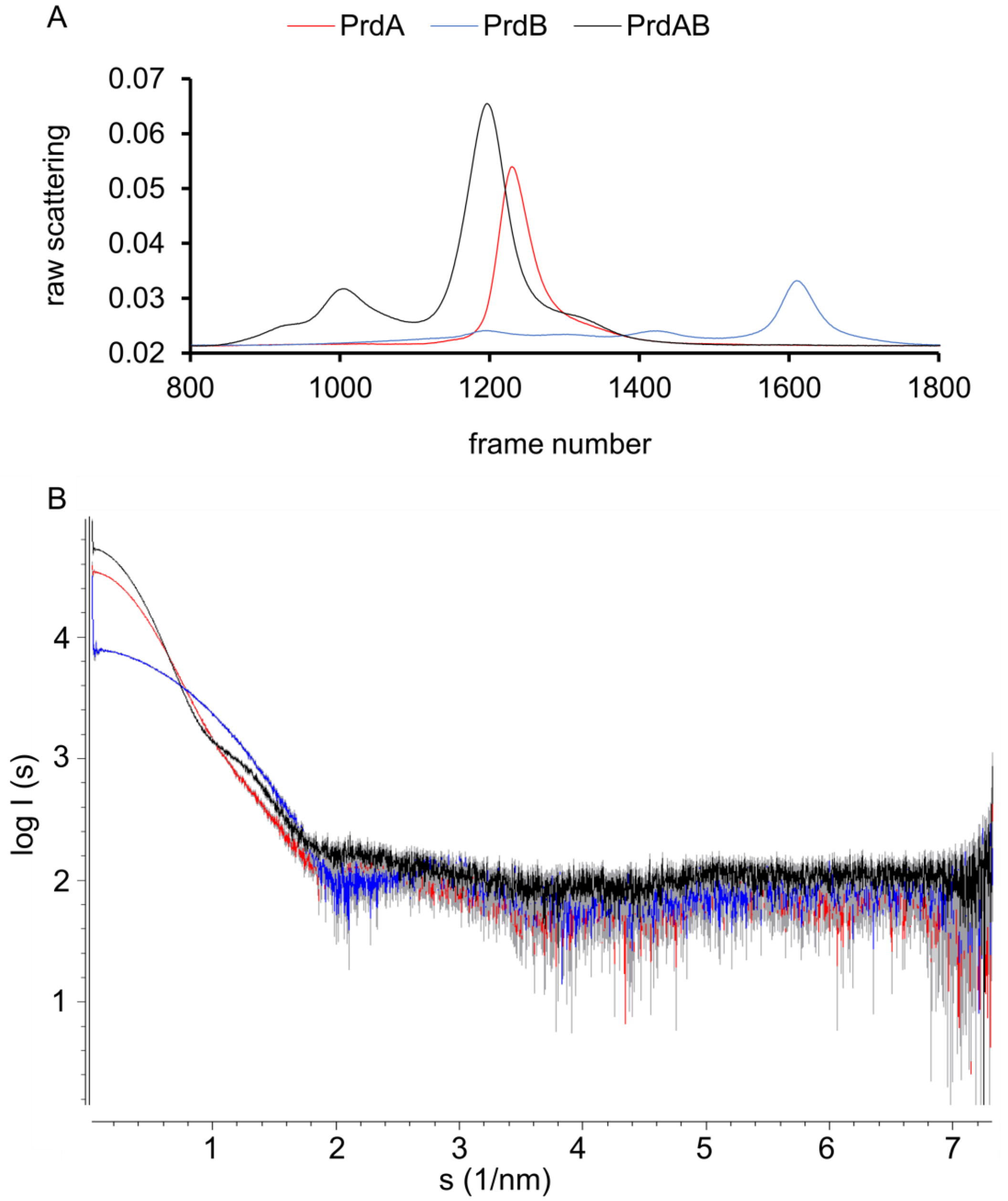


Figure 4.13: SAXS measurement on PrdA and PrdB. Experimental results of the SAXS measurement, collected at P12 at DESY, Hamburg are shown A) The raw scattering intensities for each collected frame are shown for PrdA$_{149-626}$ (red), PrdBU151C (blue) and in vitro mixed PrdA$_{149-626}$PrdB$_{U151C}$ (black) complex. A set of frames were defined as sample and background scattering. Scattering profiles were merged and subtracted. B) Merged and subtracted scattering profiles for PrdA$_{149-626}$ (red), PrdB$_{U151C}$ (blue) and the complex of PrdA$_{149-626}$PrdB$_{U151C}$ (black). The scattering profiles proof a non-aggregated condition and were used for further analysis to determine particle specific values.

The elution profile given in figure 4.13 shows that SEC separation was necessary to achieve high quality SAXS data. This way, aggregated material was removed and frames for further analysis could be selected. Merged and subtracted scattering profiles prove high data quality with significant signal up to relatively high scattering angles. As the curvature of the scattering profiles in the lowest angles up to s = 1 is convex, the samples were proven to be non-aggregated. Further, the interpolated values for I_0 correlate with the MW of the sample. The results indicate an interaction of $PrdB_{U151C}$ with $PrdA_{149\text{-}626}$, as the scattering intensity at low scattering angles increases for $PrdA_{149\text{-}626}PrdB_{U151C}$. The ATSAS package (Version 2.8.4, 2018) was used for further analysis of particle specific characters. The data range selected for analysis was chosen with scattering angles as low as possible on the lower limit. At the higher scattering angle limit, data were used until the sRg reached a value of 1.3. In this range, the linear Guinier approximation is valid.

Table 4.4: SEC-SAXS/MALS results for PrdA and PrdB. The measurement was collected at P12 at DESY, Hamburg. I_0 and R_g were interpolated from the Guinier plot. Dmax and the Porod volume are determined from the distance distribution analysis. The MW_{SAXS} is estimated by I_0, the Porod volume and further scattering parameters. The MW_{MALS} was estimated from a parallel MALS experiment.

Sample	I_0	R_g	D_{max} (nm)	Porod Volume (nm^3)	MW_{SAXS} range (kDa)	credibility MW_{SAXS} (%)	MW_{MALS} (kDa)
$PrdA_{149\text{-}626}$	34491.5 ± 12.4	3.6 ± 0.2	12.2	155.5	92.8 - 111.3	99.2	96.1 ± 0.1%
$PrdB_{U151C}$	7872.5 ± 5.8	2.0 ± 0.3	6.2	45.7	17.8 - 20.3	98.0	27.1 ± 0.4%
$PrdA_{149\text{-}626}PrdB_{U151C}$	53753.4 ± 17.9	4.0 ± 0.3	14.6	229.3	127. 5 - 151. 5	95.0	143.1 ± 0.1%

The generation of *ab initio* 3D bead-models was performed by DAMMIF. Relevant parameters were chosen as shown in table 4.4. Several modelling approaches were started and compared. One important entity for data quality assessment is the Chi^2 value. A value close to 1.0 indicates a good fit of the model to the experimental data. Models were generally generated by several repetitions. For $PrdB_{U151C}$ P1 symmetry was applied, as the protein is present as a monomer. As $PrdA_{149\text{-}626}$ was found to be a dimer, P2 symmetry was applied. For the complex of $PrdA_{149\text{-}626}PrdB_{U151C}$, P1 and P2 symmetries were tested and compared. Generated models were averaged with DAMAVER and refined with DAMMIN. The Chi^2 values were 1.083 for $PrdA_{149\text{-}626}$, 1.044 for $PrdB_{U151C}$ and 1.242 for the $PrdA_{149\text{-}626}PrdB_{U151C}$ complex. From the increase of I_0, the interaction of $PrdA_{149\text{-}626}$ with $PrdB_{U151C}$ becomes intuitively visible. Calculated values such as the MW underline this observation. Finally, the *ab initio* bead model suggests a P2 symmetry-related interaction of the subunits, were $PrdA_{149\text{-}626}$ provides a basis as a stable dimer with two distinct interaction sites to $PrdB_{U151C}$. Hence, the resulting complex most likely possesses 2:2 stoichiometry.

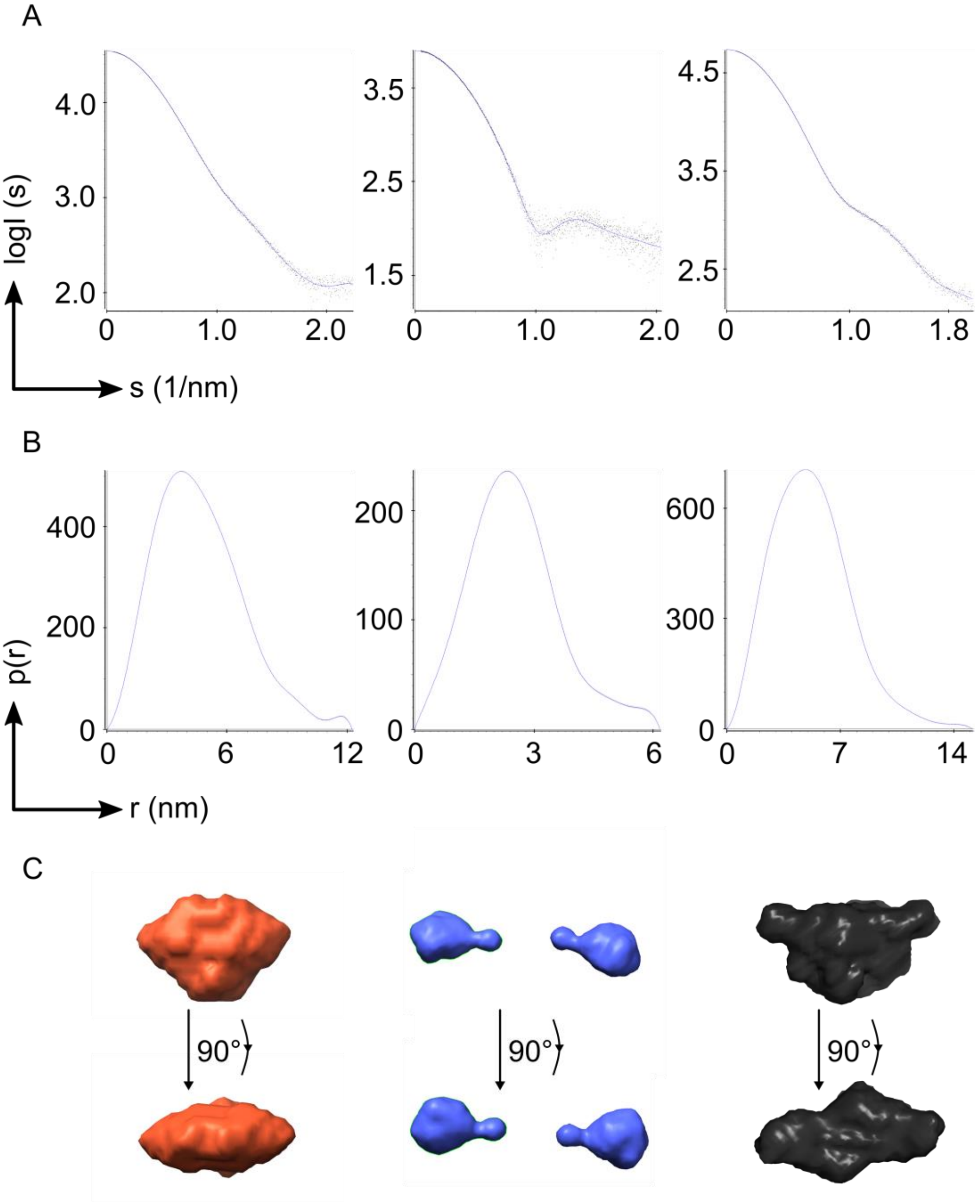


Figure 4.14: SEC-SAXS-Envelopes of PrdA and PrdB. Comparative SEC-SAXS experiments were performed with PrdB$_{U151C}$, PrdA$_{149-626}$ and the PrdA$_{149-626}$PrdB$_{U151C}$ complex. A) The scattering profiles show the data range that have been applied for 3D ab initio modelling for PrdA$_{149-626}$ (left), PrdB$_{U151C}$ (middle) and the PrdA$_{149-626}$PrdB$_{U151C}$ complex (right). B) The distance distributions for analyzed species are given here. The shape of the distance distribution curve provides first ideas of the overall shape of the particles. The distance distribution for PrdA$_{149-626}$ (left) points towards an overall globular shape. PrdB$_{U151C}$ (middle) is much smaller than PrdA$_{149-626}$ and is almost globular with an extension. The PrdA$_{149-626}$PrdB$_{U151C}$ complex (right) shows a similar shape but is elongated in one direction, as the distance distribution runs out slowly. C) Merged and refined 3D ab initio bead models are shown in the surface mode. PrdA$_{149-626}$ (red, left) has an irregular hexagonal shape. PrdB$_{U151C}$ (blue, middle) is described as a monomeric globular protein with a single extension. The complex PrdA$_{149-626}$PrdB$_{U151C}$ (right, black) shows similarity to PrdA$_{149-626}$ but it extended on two sides by appendices.

4.3.3 Cryo-EM with PrdA$_{149\text{-}626}$/PrdB$_{U151C}$

The analysis of the complex PrdA$_{149\text{-}626}$PrdB$_{U151C}$ by cryo-EM was performed by Deniz Urgular at the ThermoFisher Scientific laboratory in Eindoven, The Netherlands. PrdA$_{149\text{-}626}$ and PrdB$_{U151C}$ were purified separately and mixed in vitro as described before. The final concentration of PrdA$_{149\text{-}626}$B$_{U151C}$ was 3.7 mg/mL. The samples were aliquoted, flash frozen in LN2 and shipped on dry ice and thawed for use. For screening purposes, 10 grids were prepared in dilution series from 2-fold to 40-fold dilution. The blotting procedure was performed with a Vitrobot at 4°C and a humidity of 95%. The blotting was optimized to be carried out with 6-7 seconds and blotting force 2. A first screening was performed at a Glacios 200kV electron microscope equipped with a Falcon 3 EC camera. The magnification was set to 150,000 x. 554 images were collected with a defocus of -2.5, -2, -1.5 and -1 nm from several grids. With 1:40 diluted sample, particles were separated well enough to start data acquisition on a 300 kV Titan Krios G4 microscope, equipped with a Falcon 4 camera and automated EPU workflow. The magnification was again set to 150,000x and defocus was set to -2, -1.7, -1.4- -1.1 and -0.8 nm. Here 4133 micrographs were collected. Motion correction was applied and 405 images were sorted out due to a pixel drift of more than 50 px. Another 116 micrographs were sorted out after Ctf correction, as the resolution cut off was set to 4 Å. From 3612 micrographs, 1,497,349 particles were extracted and 985,620 were classified into only three 2D classes. These data were already sufficient to generate an ab initio model with 15 Å resolution. However, the particles were present in a single preferred orientation and hence the 2D classes showed the protein from one side only. Therefore, a second dataset was generated from a 1:160 dilution for acquisition of a dataset with a tilted specimen holder. The goniometer was tilted by 30° and another 1332 images were collected, from which 1323 remained after motion correction. Ctf estimation excluded 54 micrographs with a resolution < 4 Å. From the remaining 1269 micrographs, 570,490 particles were extracted and 426,695 of them were classified in eight 2D classes. Merging the data, further 2D classification and 3D refinement led to a 4.1 Å map from 872,193 particles. The Ctf estimation was refined and after masking and sharpening the map, a final map with 3.75 Å resolution was obtained (estimated by FSC1/2). However, the data shows a high degree of anisotropy, as the tilted dataset suffered from lower grid quality. As a consequence, merged data did not show significant improvements in map quality and density interpretation was not possible. Data processing is schematically summarized in figure 4.15. Results are shown in figure 4.16.

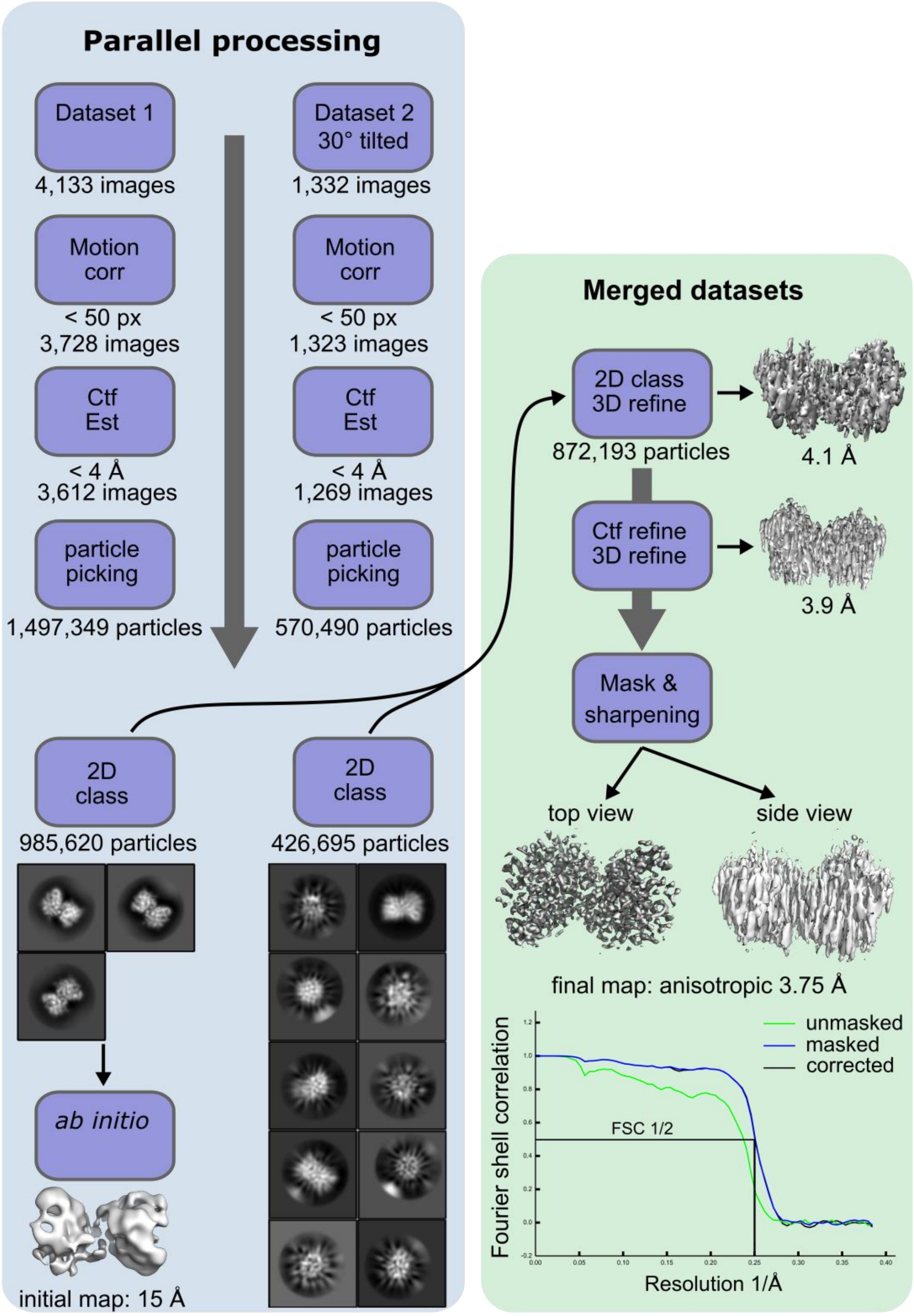


Figure 4.15: Cryo-EM workflow. The analysis of $\text{PrdA}_{149\text{-}626}\text{PrdB}_{\text{U151C}}$ hetero tetramer by cryo-EM is shown. Datasets were collected with a Titan Krios Falcon 4 and processed with Relion. Nominal magnification was set to 150,000x. Pixel size is 0.65 Å. The dose rate is 4.35 e/px/sec with an exposure time of t_{exp} = 4.86 sec. Motion correction filtered the images with a cutoff of 50 px. Ctf-estimation was used to discard images with resolution worse than 4 Å. Automated particle picking was done by EPU. 2D classification showed preferred orientation of the particles, leading to anisotropy. An initial model was generated ab initio with a resolution of 15 Å. Anisotropy was meant to be addressed by collection of a tilted dataset. However, merging data was not sufficient to eliminate anisotropy. The final map has a spatial resolution of 3.75 Å along a two-fold rotation axis (estimated from FSC-1/2-value). The resolution in an orientation perpendicular to the rotation axis is lower.

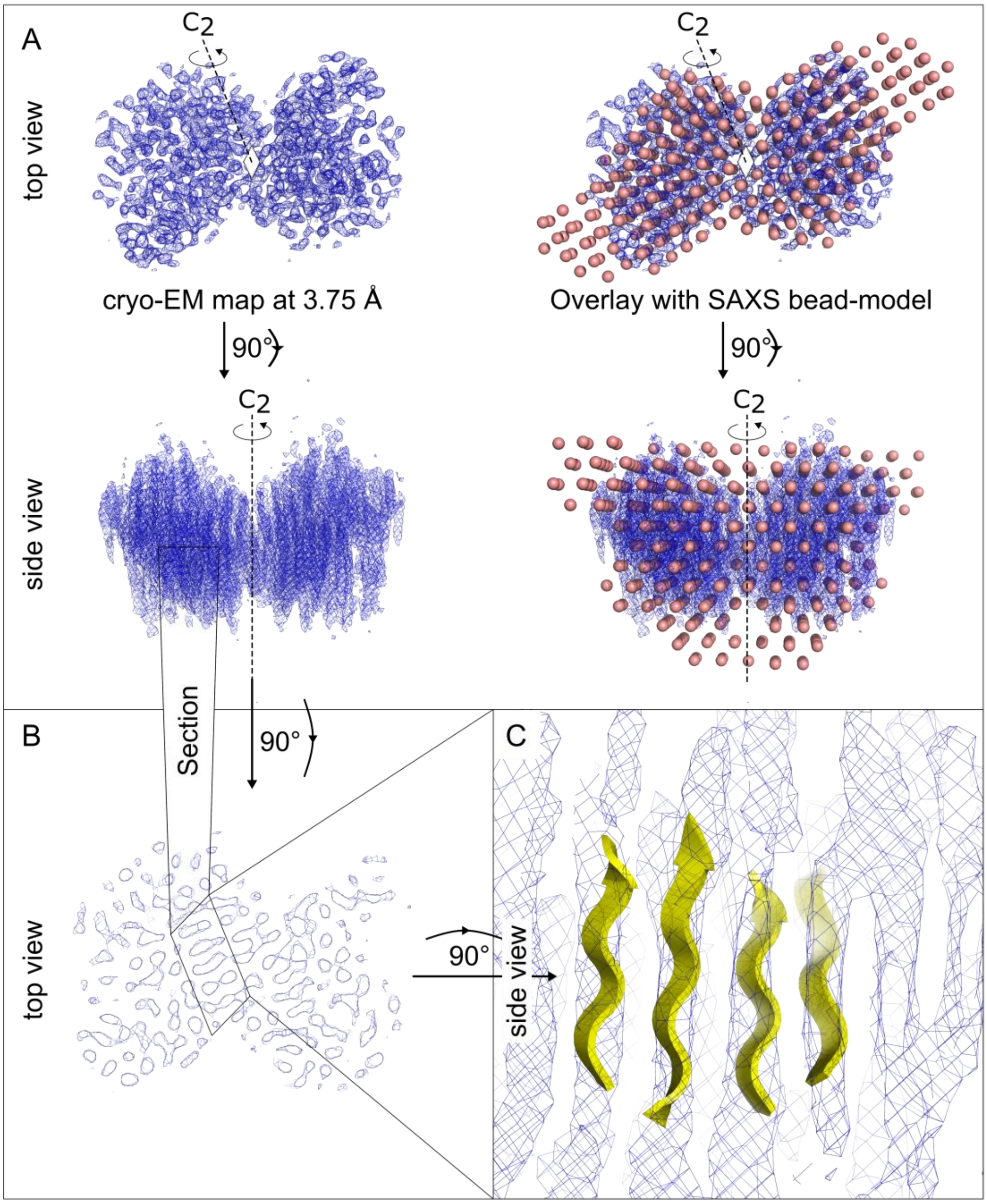


Figure 4.16: Cryo-EM results. The map for PrdA$_{149\text{-}626}$PrdB$_{U151C}$, generated by Deniz Urgular (Thermo Fisher Scientific) after merging regular data with tilted data (compare figure 4.15) is shown. A) The map revealed a two-fold symmetry with a C_2 rotation axis along the top view. The side view illustrates data anisotropy with a lower resolution due to preferred orientation of the molecules. The SAXS bead model of PrdA$_{149\text{-}626}$PrdB$_{U151C}$ (see chapter 4.3.2) suggests a two-fold symmetry. B) A section from the top view showed density that was arranged in a linear manner, characteristically found for β-sheets. However, strong anisotropy did not allow clear statements of further structural elements or atomic details. C) The described pattern could be interpreted as a set of at least 4 β-strands, arranged in a parallel sheet. However, data anisotropy impedes accurate estimations of secondary structure composition.

Direct comparison of the cryo-EM data to the results of SAXS measurements described above confirmed findings with respect to an applicable C_2-symmetry of the PrdA$_{149\text{-}626}$PrdB$_{U151C}$ hetero-tetramer. Although the map quality was insufficient to detect the symmetry automatically by cryo-EM validation tools in Phenix (Afonine et al., 2018), visual inspection allowed this interpretation (see figure 4.16A). In direct comparison to the SAXS bead model, the intact interaction of $\text{PrdB}_{\text{U151C}}$ and $\text{PrdA}_{\text{149-626}}$ could be questioned. Comparing envelope shapes for the $\text{PrdA}_{\text{149-626}}$ dimer to the $\text{PrdA}_{\text{149-626}}\text{PrdB}_{\text{U151C}}$ tetramer in figure 4.14 suggests that the cryo EM map eventually shows a $\text{PrdA}_{\text{149-626}}$ dimer. However, the data quality does not allow definite statements. Reinvestigation of the proline reductase subcomplex $\text{PrdA}_{\text{149-626}}\text{PrdB}_{\text{U151C}}$ and its smaller subunits will hopefully lead to improved map quality and deliver deeper understanding of atomic details for this so far unknown complex. Further experiments will cover the analysis of the $\text{PrdA}_{\text{149-626}}$ dimer, reanalysis of the $\text{PrdA}_{\text{149-626}}\text{PrdB}_{\text{U151C}}$ tetramer and direct comparison to the same samples after crosslinking with DSSO, which should decrease the flexibility in the subcomplex further (Kastner et al., 2008; Stark, 2010; Adamus et al., 2019). Data anisotropy, resulting from only one preferred orientation in the EM grid, clearly represents the most important hurdle in sample preparation of the proline reductase subcomplex. It is hypothesized that this common phenomenon is caused by the thin layer of solvent, leading to molecular adsorption at the air-water interface (Chen et al., 2019). Collecting tilted datasets as it has been performed here, is only one of several means to overcome data anisotropy without applying changes to the environment of the sample. This technique is applied on a regular basis and well established (Zi Tan et al., 2017; Leber et al., 2018; Ghanim et al., 2019). However, in this study this approach was not successful, as reproducibility of high grid quality was not given. Another approach would be sample treatment that leads to a more randomized orientation. Examples are the addition of zwitter-ionic detergents, such as CHAPSO, or the use of grids with modified surfaces (Drulyte et al., 2018; Chen et al., 2019; Wang et al., 2020). In conclusion, cryo-EM seems to be a valuable method to elucidate the architecture of *C. difficiles* proline reductase subcomplexes, although data finalization and map interpretation was not possible at the current state of research.

4.4 Proline reductase complex natively isolated from *Clostridioides difficile*

The complete and untruncated complex of PrdAB was not accessible by recombinant protein production. Possibly, a Gram-positive expression system, such as *Bacillus megaterium*, could circumvent these difficulties and produce proteinogenic material with correct fold and maturation. Nevertheless, as this complex is unique for *Clostridia*, it remains unclear whether the native source supplies a specific chaperon system for correct complex assembly.

Consequently, the complex of PrdAB had to be purified from its native source. In this study, *Clostridioides difficile* 630Δerm was cultivated in a proline-containing CDMM and harvested in the late exponential growth phase.

4.4.1 Optimization of the existing purification protocol for the proline reductase complex

The purification procedure for the proline reductase is based on the work of Kabisch et al. (Kabisch et al., 1999). As the protein complex was to be used for structural investigations and further biophysical measurements, an as gentle as possible protocol was sought. In order to achieve this, several changes had been applied to the previously published protocol to reduce the number of purification steps and to avoid an $(NH_4)_2SO_4$ precipitation.

One pellet from 1.6 L culture was resuspended in 40 mL lysis buffer. Upscaling is not recommended, as the following AC step does not have sufficient capacity to bind the entity of material, the proline reductase will consequently be removed from the column competitively. For direct precipitation of poorly soluble proteins immediately after cell lysis, 40 mL of cold 2 M $(NH_4)_2SO_4$ solution was added slowly under stirring, leaving the proline reductase complex in the soluble phase. Cell debris were removed.

Hydrophobic interaction chromatography (HIC) was optimized with a HIC selection kit, identifying butyl sepharose (Butyl HP) as the affinity material with the highest specificity for the proline reductase complex.

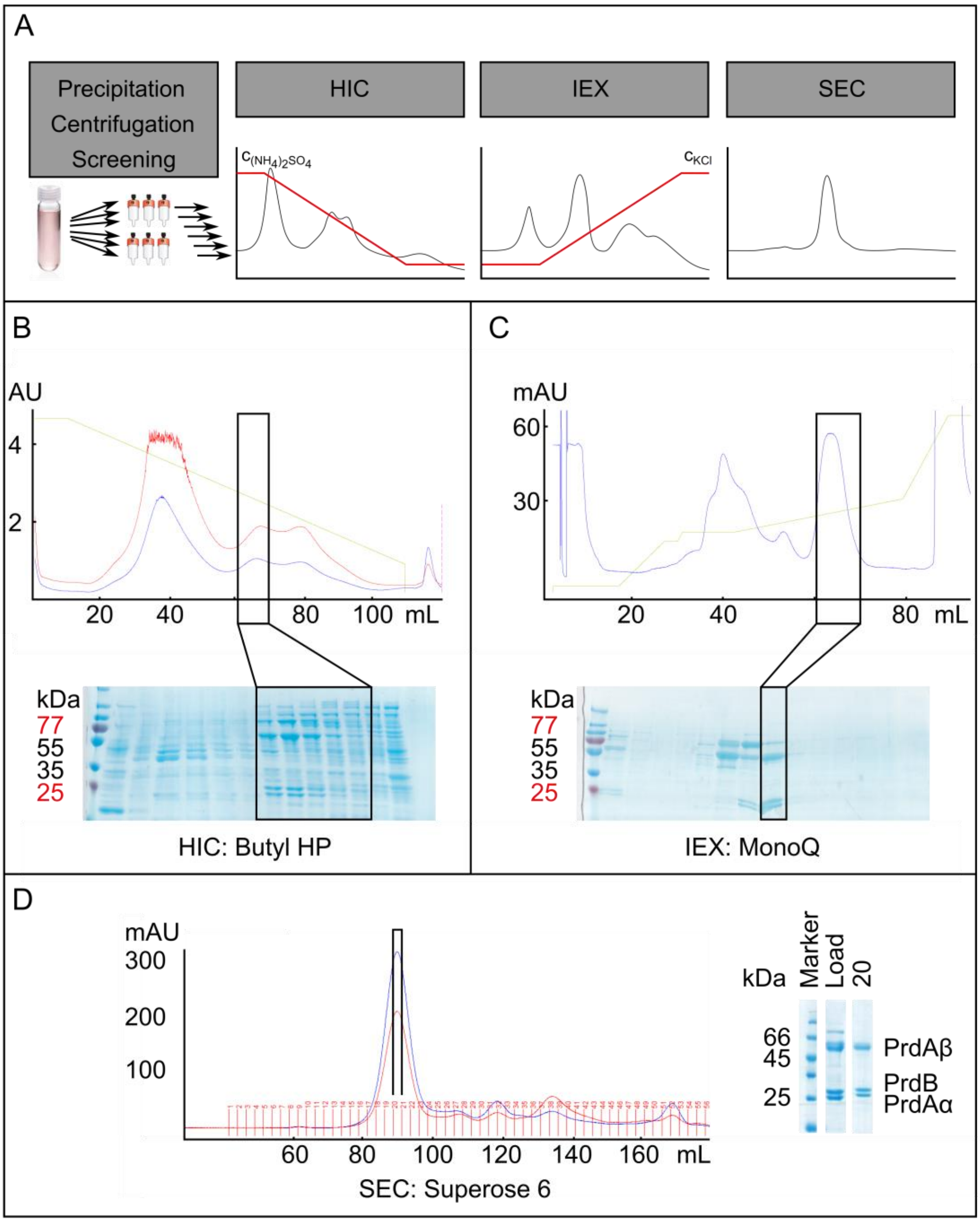


Figure 4.17: Summary improved purification of native proline reductase complex. A) Schematic workflow. Lysed cells were mixed with 1 M $(NH_4)_2SO_4$ for precipitation of poorly soluble proteins and spun down by high performance centrifugation. The supernatant was split and applied to six different HIC materials for screening purposes. IEX chromatography was applied as an intermediate purification step. The target is polished by SEC. B) The cleared lysate was applied on a 5 mL Butyl HP column. The $(NH_4)_2SO_4$ concentration was reduced from 1 M to 0 M. Eluting material was analyzed by SDS-PAGE and pooled as indicated by the black box. C) A fine-tuned KCL gradient was applied in the intermediate IEX step. Different species of proline reductase containing fractions were obtained and analyzed by SDS-PAGE. The latter species was pooled and kept for further purification. D) HIC-IEX-isolated proline reductase complex was polished by SEC. Last impurities were removed. SDS-PAGE shows three bands, corresponding to PrdAβ, PrdB and PrdAα, as confirmed by mass spectrometry.

The IEX purification step, using MonoQ material, was optimized solely in salt gradient slope and step arrangements. An additional band was observed after IEX purification on the SDS-PAGE. This protein was identified to be the elongation factor G FusA. A direct interaction of FusA could not be confirmed by means of MST and TSA. Sequence based calculation of the theoretical pI shows a value of 4.91, which is close to the value for the proline reductase complex of 4.83. Hence, this protein might be an impurity that coelutes with the target. The final purification step was a SEC, using a Superose 6 16/100, equilibrated with PR SEC-MALS buffer, separating the PR complex from FusA. Results are summarized in figure 4.17.

4.4.2 Thermal shift assay with the proline reductase complex

In order to obtain robust protein complexes for TEM and cryo-EM, the buffer conditions for the proline reductase complex were revised. TSA provides a powerful tool to determine melting temperatures of proteins in a 96-well format. Therefore, purified proline reductase complex was addressed by TSA.

In the first attempt, the pH optimum was determined as described in chapter 3.6.2. The calculated melting temperature was plotted against the pH value, as shown in figure 4.18.

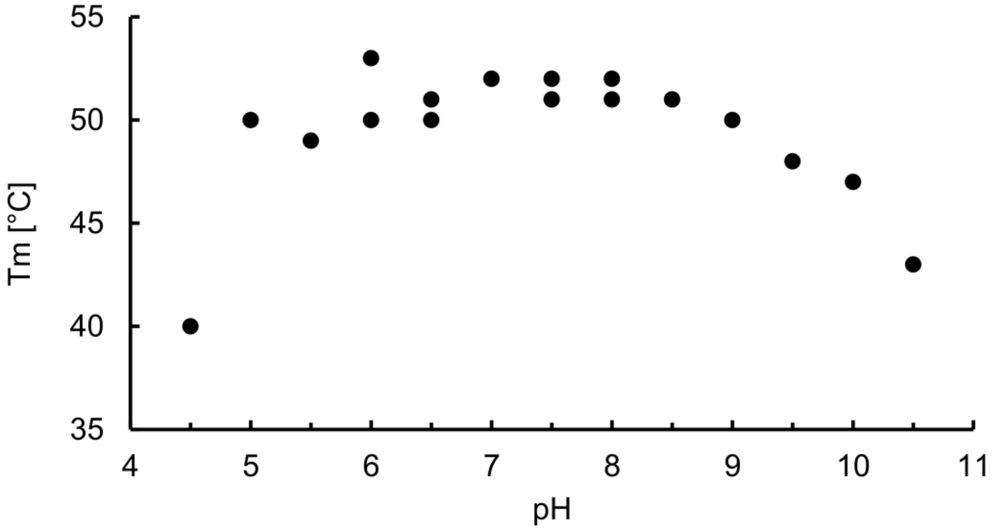


Figure 4.18: Determination pH optimum for the proline reductase complex. The TSA buffer screen was suitable to determine a pH dependent melting temperature. A range from pH 4 to pH 11.5 was tested by mixing the purified complex with different buffer substances. As a result, the ph range from 7 to 8 was found as a stability maximum with a T_m of 54°C.

The determined pH optimum was in the range of 7 to 8.5.

Further, additive screening was performed to identify compounds that contribute to protein complex stability. No compounds could be identified that further increase the melting temperature. Addition of TCEP leads to a complete loss of melting signal. As a similar effect on PrdA$_{149\text{-}626}$ was observed for EDTA and EGTA (compare chapter 4.1.5), it was contemplated that chelating agents affect complex stability. In fact, chemical features of TCEP do not only provide reducing properties, but also enable the molecule to act as a chelator. Indeed, studying the effect of EDTA and EGTA on the full proline reductase complex shows a significant decrease of melting signal compared to KCl and DTT.

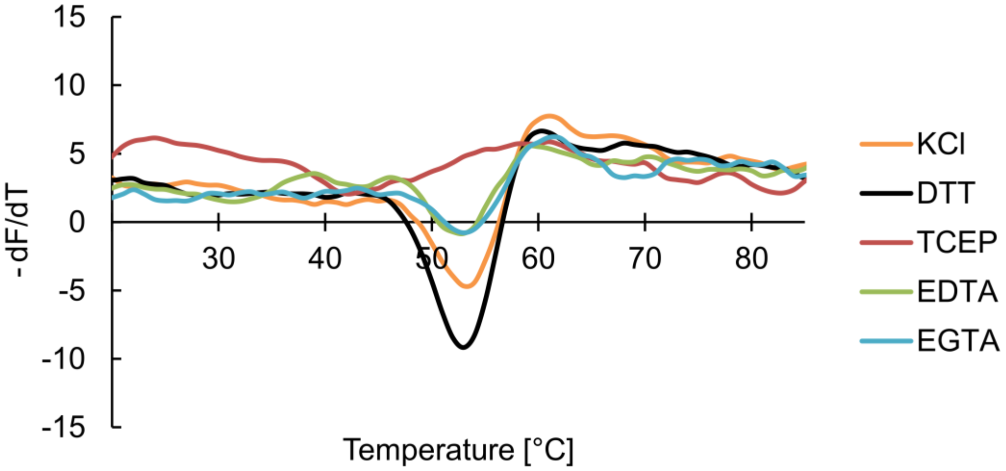


Figure 4.19: Melting profiles for proline reductase complex, chelators. Melting profile of the proline reductase complex. Addition of KCl and DTT does not affect protein stability. However, the supercomplex integrity is negatively affected by the addition of chelators, such as TCEP, EDTA and EGTA.

Possibly, the proline reductase complex assembly is mediated by metal ions. Similar observations have been described before for various ions, such as zinc, copper or manganese (Stott et al., 1995; Sharp, 1996; Maret, 2012; Pace et al., 2014; Song et al., 2014; Garapati et al., 2018). Hence, addition of strong chelators must be avoided in purification. Purification protocols were adjusted by using EDTA-free protease inhibitors.

4.4.3 GraFix and Blue native-PAGE

The MW of the proline reductase complex has previously been estimated by Kabisch et al. using analytical SEC and a standard MW calibration (Kabisch et al., 1999). It was described as an 870 kDa complex, possibly representing a nonamer or decamer of trimers, which is in line with observations made in this study. However, as the MW estimation by SEC is dependent on the hydrodynamic radius and therefore on the shape of the particle, it usually provides relatively rough estimates only. In order to understand the architecture of the complex, an exact determination of the stoichiometry is required, and techniques that are more accurate were applied here.

An alternative technique for MW determination is the Blue native (BN)-PAGE. This electrophoretic separation method does not contain denaturing or reducing agents. Instead, the charge required for mobility in the electric field is generated by mixing the protein with the anionic dye Coomassie blue G250. In many cases, Coomassie attaches to the surface of proteins by ionic interaction to arginines without inducing unfolding, hence allowing uniform mobility in an electric field. Here BN-PAGE was combined with the GraFix approach (Kastner et al., 2008; Stark, 2010) to stabilize the delicate complex. Here, this separation technique was established without crosslinkers and compared to results were the maximum concentration of glutar aldehyde (GA) as a crosslinker was set to 0.01 and 0.1%. Elution profiles show high background absorption, which results from GA (compare figure 4.20 A). After the GraFix run, non-crosslinked proline reductase complex was analyzed by SDS-PAGE (figure 4.20 B). Crosslinked proline reductase complex was applied to the BN-PAGE with a separation range suitable for large complexes (figure 4.20 C).

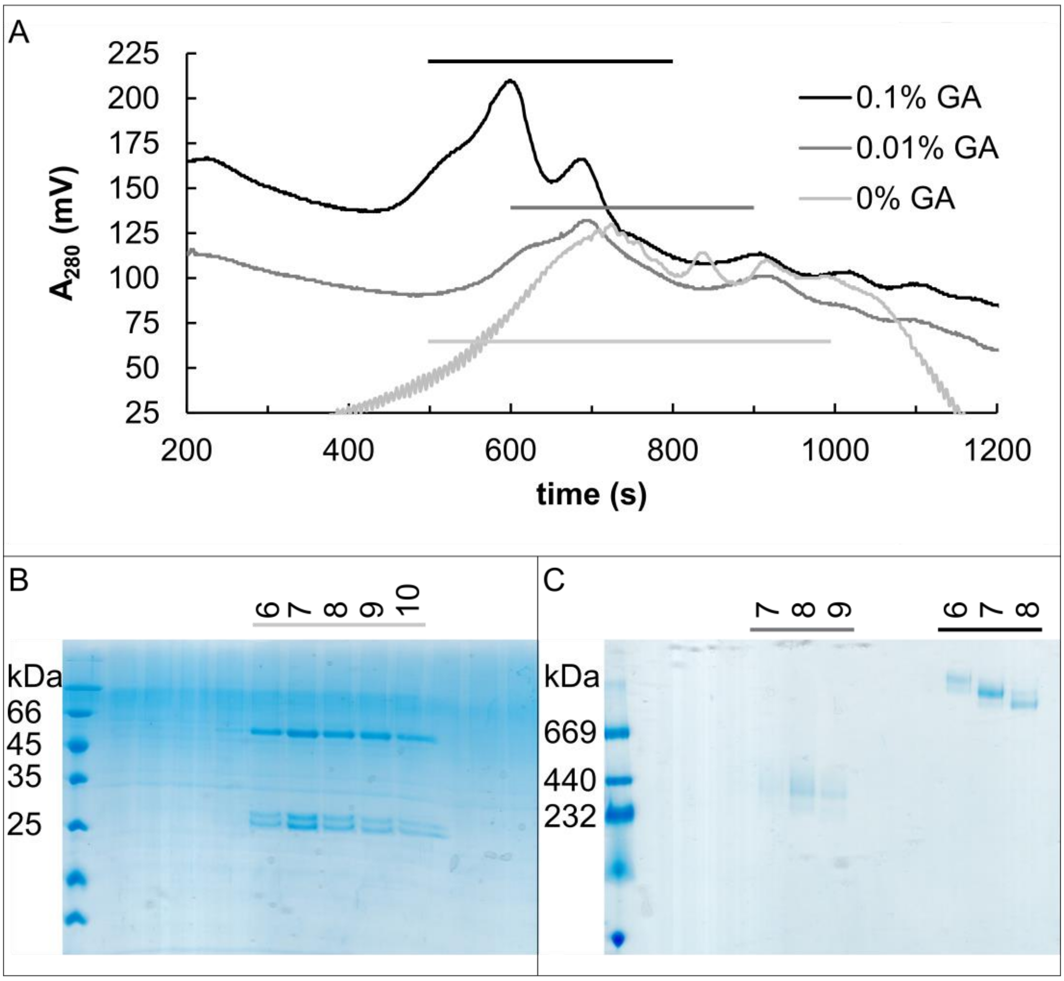


Figure 4.20: GraFix and BN-PAGE results for proline reductase complex. The complex was purified from *C. difficile* and analyzed in a native and crosslinked state. A) Comparative elution profiles of proline reductase supercomplex after gradient ultracentrifugation separation in a glycerol gradient from 15 – 35% glycerol. Applying 0.1% GA in the 35% glycerol phase results in three eluting species, from which two lie under one peak eluting after 500 seconds. The third species elutes later after almost 700 seconds (black). Decreasing the GA concentration by a factor of 10 leads to a retarded elution after 600 seconds. A broad peak is observed (grey). The control run without crosslinker shows a similar elution profile with a relatively broad peak and several eluting species (light gray). Time-based fractionation was applied with fractions of 100 seconds each. Fractions were applied to SDS-PAGE for the control and to BN-PAGE for crosslinked samples as indicated by bars in the corresponding colors. B) SDS-PAGE for the control separation of the proline reductase supercomplex without crosslinker. In fraction 6 to 10, all three subunits of the complex can be found (light grey). C) The BN-PAGE for samples crosslinked by 0.01% GA (grey) and 0.1% GA (black) are shown. For 0.01% GA, the elution profile is comparable to non-crosslinked proline reductase. Here a complex of approximately 400 kDa can be observed, which would correspond to a tetramer of trimers with a theoretical mass of 372 kDa. Applying higher concentrations of 0.1% GA, the elution profile shifts towards higher MW that is larger than 669 kDa. Precise MW determination is not possible here. The proline reductase complex forms three species with different migration depths, corresponding to three eluting species monitored in section A).

The elution profiles recorded a shift in apparent MW for increasing concentrations of GA (see figure 4.20 A). Inspecting the SDS-PAGE for the control GraFix run in figure 4.20 B, one can see a good performance of separating proteins based on their size and a high purity of the proline reductase subcomplex. Supercomplex integrity cannot be judged from this SDS-PAGE, but the complex of PrdAα, PrdAβ and PrdB is maintained. Comparing the elution profiles and gel patterns (Figure 4.20A and B) to those of crosslinked proline reductase complex (Figure 4.0 C), one can see a shift of MW from approximately 400 kDa to more than 800 kDa by increasing the GA concentration to 0.1%. The lower MW species of around 400 kDa is in line with early findings from Seto and Stadtman (Seto et al., 1976). However, this GA concentration-dependent mass-shift in combination with our previous findings, reported in this thesis, indicate that BN-PAGE is not an appropriate method to determine the MW, as the complex is easily losing its quaternary structure. It was shown in previous experiments that truncated $PrdA_{149\text{-}626}$ is able to form stable dimers, while PrdB is present as a monomer (see also chapter 4.3). However, on the BN-PAGE these lower MW species of 90 kDa or 25 kDa were not observed, indicating, that although the supercomplex disintegrates during the electrophoresis, the PrdAB-subcomplex still persists. Instead, the decomposition of the PrdAB-supercomplex into PrdAB-heterohexamers seems to be the reason for observed results.

4.4.4 *In-situ* dynamic light scattering

Time-resolved *in situ* DLS experiments were performed with support from Arne Meyer (Xtal Concepts, Hamburg) and was used as a direct feedback for oligomerization, aggregation and radii calculations. The direct readout of particle conditions in various environments gives information on potential improvements in negative stain EM. For an accurate measurement, five repetitions at different concentrations ranging from 10 mg/mL to 0.5 mg/mL of the proline reductase supercomplex were performed. The radius was determined to be 8.0 nm ± 1.3%. From this value the MW was calculated as 396.5 kDa ± 3.0%, which again correlates with early findings from Seto and Stadtman (Seto et al., 1976). Additionally, additive screening was performed to test the influence of previously identified additives on the oligomerization state directly (see chapter 4.4.2). PBS and HEPES buffers were tested in a range from pH 7 to 8 as a storage condition. Additionally, phosphor tungstate acid (PTA) as a negative stain reagent was tested in a pH range from 5 to 7 and $(NH_4)_2MoO_4$ at pH 7. Acetic acid and formic acid were tested at pH 4.3, mimicking the environment while staining with uranyl acetate (UAc) and uranyl formic acid (UFA).

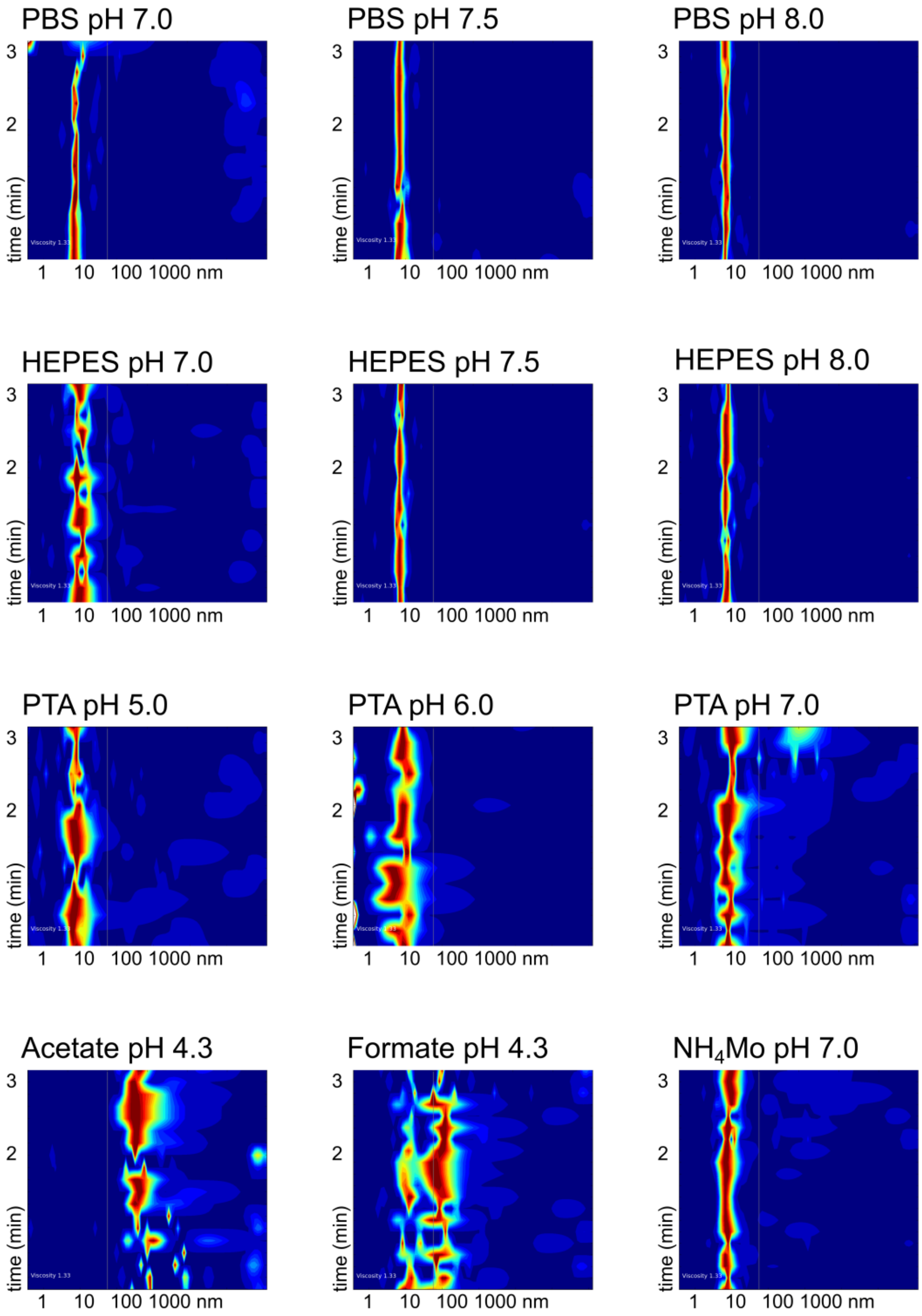


Figure 4.21: Results of in situ DLS for the proline reductase complex. The protein solution was mixed in a 1:1 ratio with different buffers (100 mM) and the scattering was observed over a period of 3 minutes. A sharp, defined signal that was stable over time was observed for buffer substances in the range of pH 7 to 8. For PTA, aggregation was observed, independent from the pH. Low pH buffers, mimicking negative stain conditions, such as acetate and formate show immediate heavy agglomeration. The dispersity for NH$_4$Mo at pH 7 appears to be more conserved, compared to other staining methods. However, it was still negatively affected when comparing the signal to pure storage buffer conditions. Calculated radii and MW estimations are given in table 4.5.

As has been observed before (chapter 4.4.2), the complex is most stable in the pH range of 7 to 8.5. In agreement to this, the DLS signal for those buffer conditions was stable over a longer period of time and showed neither disintegration, nor agglomeration for both, PBS and HEPES buffer systems. A different behavior can be seen for the sample HEPES pH 7, which might be elicited by improper sample treatment during the mixing procedure or placement under oil. In general, this observation attests high sample quality.

Table 4.5: Results in-situ DLS on proline reductase complex. The data corresponds to the measurements depicted in figure 4.21. Calculated radii and MW estimations are given for samples mixed with storage buffers and negative stain reagents.

Condition	**Radius (nm)**	**MW (kDa)**
PBS pH 7.0	9.1	528.6
PBS pH 7.5	8.4	431.7
PBS pH 8.0	8.7	472.5
HEPES pH 7.0	13.0	1185
HEPES pH 7.5	8.8	489.2
HEPES pH 8.0	9.1	527.4
PTA pH 5.0	10.0	652.0
PTA pH 6.0	10.2	674.9
PTA pH 7.0	10.9	797.0
Na-Acetate pH 4.3	aggregates	
Na-Formate pH 4.3	aggregates	
NH_4Mo pH 7	9.9	630.8

As described in chapter 4.4.2, the proline reductase complex is sensitive to low pH values. The *in situ* DLS measurements support this finding for conditions mimicking negative stain procedures, as heavy agglomeration was observed upon addition of acetic acid and formic acid at pH 4.3. This sensitivity must be considered before loading the sample onto a TEM grid. $(NH_4)_2MoO_4$ at pH 7 seems to affect protein integrity to a lesser extent. Therefore, this agent can be considered a useful staining reagent, despite its low contrast in negative stain TEM.

Testing DTT and TCEP as reducing agents, it became clear that DTT has no influence on protein stability or integrity, as shown in figure 4.22. However, addition of TCEP leads to heavy agglomeration. Again, this is in agreement with the TSA results (compare chapter 4.4.2). Due to high buffer concentrations, a pH drop is ruled out to cause this behavior. Calculated radii and MW estimations are summarized in table 5.5.

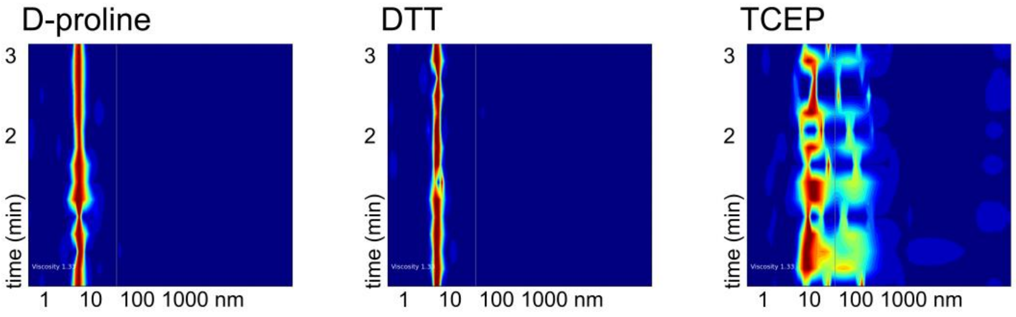


Figure 4.22: Results of in-situ DLS for the proline reductase complex. After addition of different additives. D-proline and DTT both did not show any effect on oligomerization of the complex. Addition of TCEP however, lead to immense aggregation of the proline reductase.

As investigating the architecture of the complete complex is of priority, the effect of adding crosslinking agents was monitored by *in-situ* DLS with the aim to determine crosslinking conditions that stabilize the supercomplex.

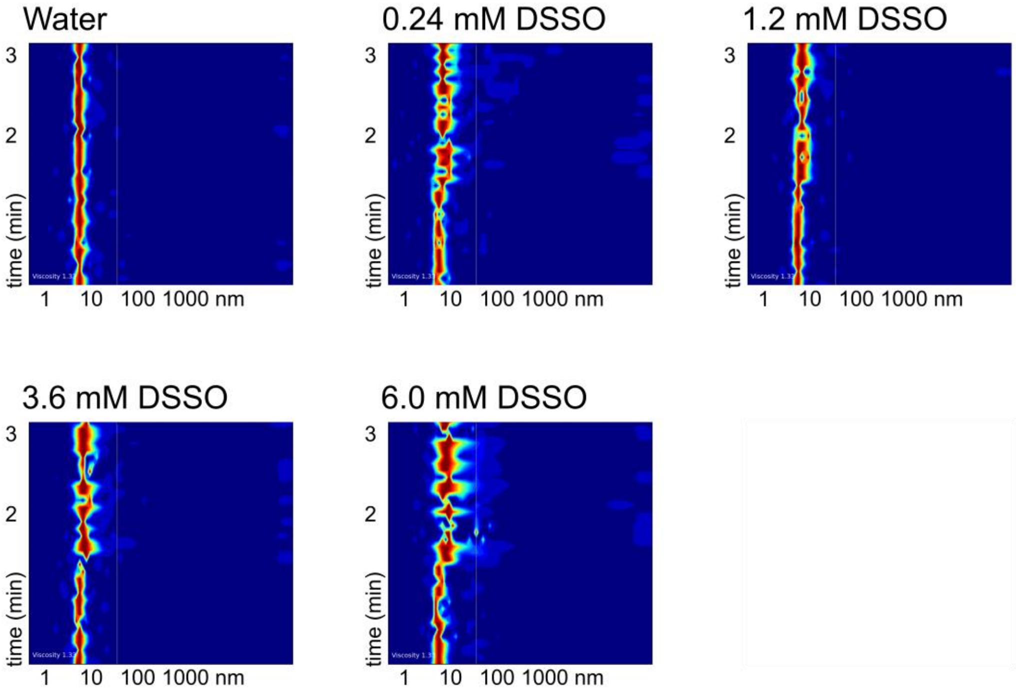


Figure 4.23: Results in situ DLS crosslinking experiments. The proline reductase complex was crosslinked with various concentrations of DSSO. The distance distribution of the proline reductase supercomplex gradually increases upon the addition of rising concentrations of the crosslinker DSSO. However, the obtained species do not seem to be monodisperse.

For this, the proline reductase was mixed with various concentrations of DSSO. The *in-situ* DLS results for crosslinked proline reductase is shown in figure 4.23.

Table 4.6: Results in-situ crosslinking experiments. Upon the addition of the crosslinker DSSO, the measured radius, as well as the estimated molecular weight gradually increases.

Condition	**Radius (nm)**	**MW (kDa)**
Control	8.4	431.6
0.24 mM DSSO	8.5	451.5
1.2 mM DSSO	9.4	571.7
3.6 mM DSSO	9.8	618.2
6.0 mM DSSO	10.5	725.1

In analogy to the results of the BN-PAGE, crosslinking was introduced prior to measurement. Figure 4.23 shows an effect on scattering behavior, but the signal does not show a clear shift. At the same time, agglomeration was not observed, pointing towards a stabilization of a higher MW complex that is potentially present in native conditions. As one can see in table 4.6, both, the calculated radii, as well as the estimated MW rose with increasing crosslinker concentration, approaching values comparable to analytical SEC. This was encouraging to confidently use crosslinkers for exact size determination and negative stain TEM blotting. As the scattering did not shift completely to a distinct higher oligomeric MW species, but a fraction remained at the previous level of 8 nm, it appears likely, that the maximum concentration of 6 mM DSSO was not sufficient to link the complete amount of protein.

4.4.5 SEC-MALS with the proline reductase complex

SEC-MALS experiments on the proline reductase complex purified from *C. difficile* were performed with a Superose 6 10/30 increase column. The analysis was performed with the ASTRA® software (Wyatt Technology Corporation, USA). The complex was analyzed in terms of size and behavior upon crosslinking. Further, SEC-MALS was used to identify eluting species that are suitable for negative stain TEM grid preparation. As SEC-MALS is a robust method to determine MW's, repetitive measurements with increasing concentration of the crosslinker DSSO were performed to determine conditions that lead to the highest level of crosslinking possible without the formation of aggregates. The first attempt of this experiment was set up with 1 mM DSSO to perform a proof of principle experiment with 280 µg for both samples to enable an exact measurement and comparability.

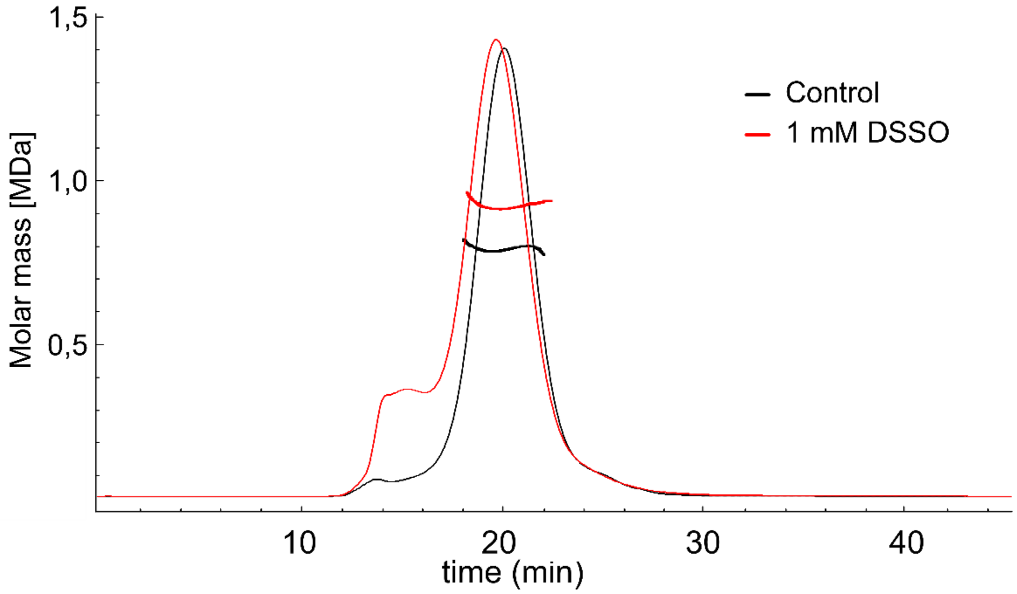


Figure 4.24: SEC-MALS comparison native and crosslinked proline reductase complex. Direct comparison of native (black) and crosslinked (red) proline reductase complex, purified from native source in a SEC-MALS experiment. Both samples show a distinct scattering peak. The native sample (black) has a calculated molecular weight of 794.2 kDa. After crosslinking the sample with 1 mM DSSO, the molecular weight shifted to 935.2 kDa. The elution volume is slightly shifted to a higher gyrical radius.

In a direct comparison of native and crosslinked proline reductase (see figure 4.24) it was found that the elution volume slightly shifted to a higher hydrodynamic radius. At the same time, the MW shifted from 794.2 kDa to 935.2 kDa.

As the crosslinker should neither alter the hydrodynamic radius that determines the elution volume from a SEC column nor the refractive index increment (dn/dc) used to calculate MW from scattering, these results point towards a stabilization of a decamer of PrdAB-hetero-trimers. Unlike BN-PAGE and *in situ* DLS, SEC appears to be a gentler method to maintain the complete quaternary structure of the proline reductase supercomplex. However, there seems to be still dynamic disintegration and assembly on the column, which leads to an elution peak with higher elution volume and lower calculated MW from MALS. As one can see from the scattering profile, aggregates that are separated from the main species are formed to some extent. Considering that scattering of large aggregates has higher intensity, the amount of aggregate is considered as low. This is supported by a negligible UV signal for the elution peak that represents these aggregates. Although the maximal level of crosslinking was probably not reached, the peak fraction of the eluted complex was collected and used for negative stain TEM. At the same time, crosslinking was repeated with higher amounts of crosslinking agent and a more economical input of proline reductase. Concentrations of DSSO tested in the range from 2 mM to 20 mM. The results are shown in figure 4.25.

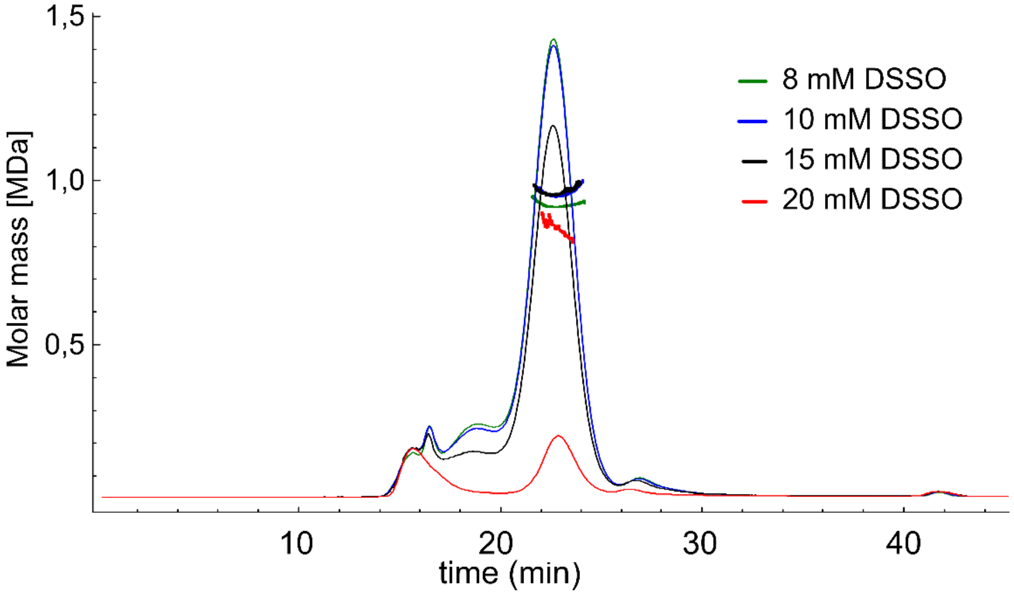


Figure 4.25: SEC-MALS results for increasing concentrations of crosslinker. The crosslinking agent DSSO (8 mM green, 10 mM blue, 15 mM black and 20 mM red) was applied to 50 µg of proline reductase complex each. The injected volumes and masses were kept constant. The elution volume did not change by increasing the crosslinker concentration. After adding 20 mM DSSO (red) to the proline reductase complex, the scattering signal was significantly decreased significantly.

By comparing elution profiles for increasing crosslinker concentrations, the effect on the calculated radii and the MW was marginal for most concentrations of DSSO tested. For DSSO concentration below 15 mM, the content of aggregates persists at a low level, while concentrations above 15 mM DSSO seem to initiate agglomeration. Results are summarized in table 4.7.

Table 4.7: Summary of SEC-MALS results for increasing concentrations of crosslinker. The proline reductase supercomplex was mixed with different concentrations of DSSO. The native complex was addressed by MALS experiments with increasing crosslinker concentration.

Sample	calc. MW (kDa)	Uncertainty (%)	Oligo factor
native	794.6	0.2	8.5
1 mM DSSO	927.0	0.0	10.0
2 mM DSSO	943.8	0.4	10.1
3 mM DSSO	973.3	0.5	10.5
4 mM DSSO	967.1	0.4	10.4
8 mM DSSO	927.7	0.4	10.0
10 mM DSSO	962.5	0.4	10.3
15 mM DSSO	967.1	0.2	10.4
20 mM DSSO	928.1	3.6	9.2
$\overline{\text{Ø}}_{\text{crosslinked}}$	940.5	3.9	10.1

The main species from the crosslinking with 15 mM DSSO was collected for negative stain TEM grid preparation. Comparisons of TEM micrographs of native and crosslinked proline reductase are given in chapter 4.4.7. Comparing the MW for native and crosslinked proline reductase supercomplex, it seems that the complex increased in size, regardless of the applied concentration of DSSO. This points towards a stabilization of a decameric supercomplex. Presumably, the supercomplex has a low stability and is present as a mixture of octamers and decamers when no crosslinker was applied. Crosslinking has been used widely for stabilization of delicate protein complexes (Schmidt et al., 2017; Rozbeský et al., 2018; Adamus et al., 2019; Chu et al., 2019)

4.4.6 SEC-SAXS with the proline reductase complex

In analogy to SEC-MALS experiments, the proline reductase supercomplex was investigated with SEC-SAXS. SAXS data was collected at beamline P12 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). Beamline operation was supervised by Cy Jeffries (Blanchet et al., 2015). The supercomplex was crosslinked to support the complete quaternary structure, as dynamic arrangement and disruption of the supercomplex was derived from experiments performed by crosslink SEC-MALS. As expected, the scattering profile shows similarities to previous results with high MW aggregates, as shown in figure 4.26. Data analysis was performed with the ATSAS package (Franke et al., 2017). Scattering images for merging were selected based on a constant value for R_g, which was calculated with 9.8 nm to 9.6 nm over the images from 1189 to 1268. The background selection was done automatically by CHROMIXS. The scattering profile resulting from the subtraction of the background from the peak data was used for further analysis. Data analysis was used to calculate the MW, as Guinier analysis allows at low scattering angles from the scattering profile (Mylonas et al., 2007). The Rg was calculated in the Guinier range for the data points from 10 to 40. I0 was interpolated

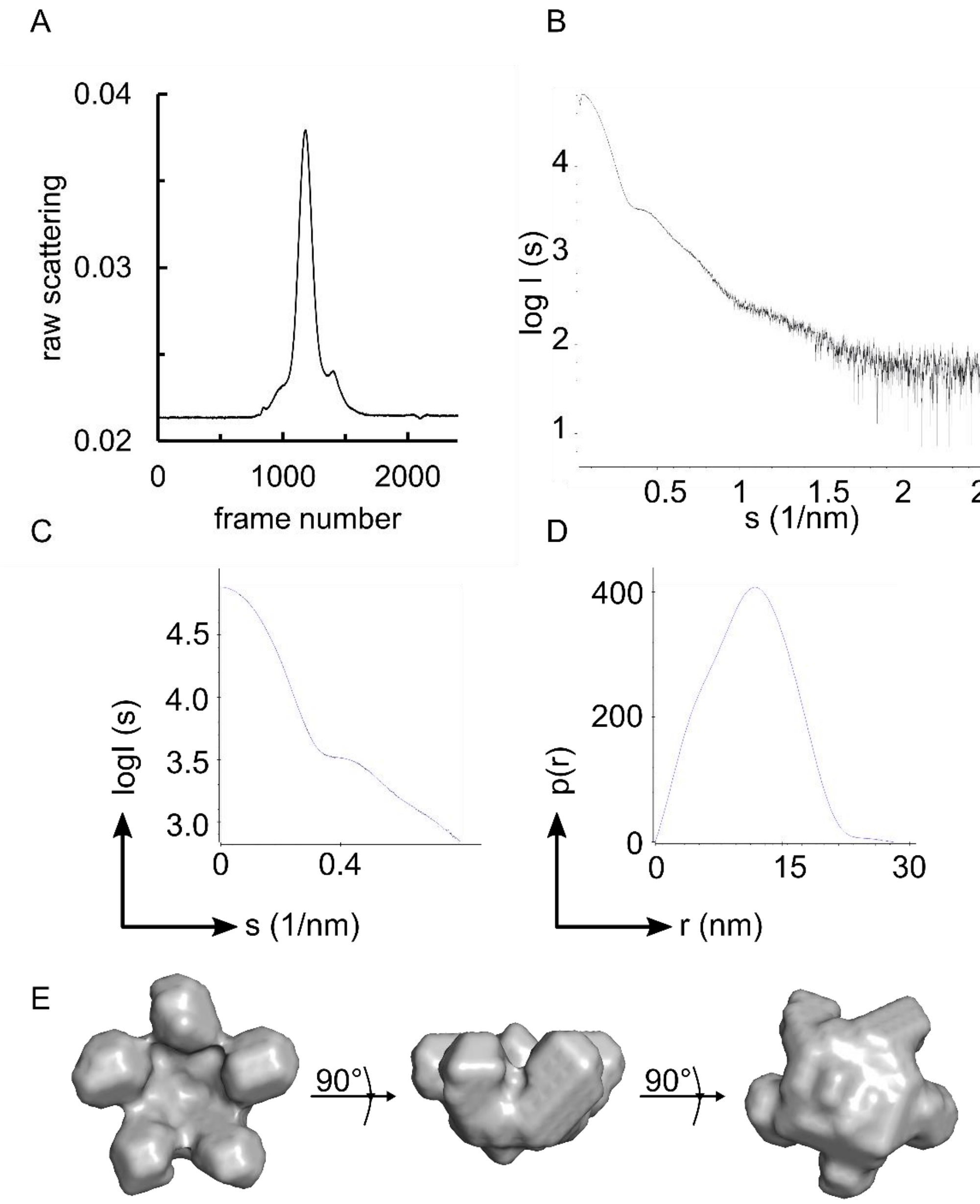


Figure 4.26: Results SEC-SAXS for the proline reductase complex. The results are collected on natively purified and crosslinked proline reductase supercomplex. A) The raw scattering over frame number constitutes the elution profile of the complex. A sharp peak was separated from minor portions of aggregates and disintegrated complex. The background was defined in an early area of the elution. B) Subtraction of sample and background scattering curves resulted in this scattering profile. From the scattering intensity at low angles, I_0 was derived. The versatile features of the curve point towards a complicated shape. C) For 3D bead model generation, scattering data was used to a certain scattering angle, as depicted here. The calculated curve fits the observed data to a high degree. D) The distance distribution unveils a number of features that were used to calculate a 3D bead model with a D_{max} of around 32.5 nm. E) The 3D bead model generated by assuming a 5-fold symmetry is shown. It was calculated with a Chi^2 value of 1.183.

with 76656.4 ± 69.2. Rg was calculated from the slope to be 10.01 ± 0.58 nm. This way the sRg limits were adjusted to 0.49 to 1.3. The calculated MW was 829.5 kDa to 1237.5 kDa with a credibility of 92%. In comparison, the MW was determined by MALS with 880 kDa (direct comparison to 940.5 kDa in independent SEC-MALS experiment). Results are summarized in table 4.8.

Table 4.8: Results SAXS/MALS measurements for the proline reductase complex. The data is collected at beamline P12 at DESY, Hamburg. I_0 and R_g were interpolated from the Guinier plot. Dmax and the Porod volume were determined with the distance distribution analysis. The MW_{SAXS} was estimated by means of I_0, the Porod volume and further scattering parameters. The MW_{MALS} was estimated from a parallel MALS experiment.

Sample	I_0	R_g (nm)	D_{max} (nm)	Porod Volume (nm^3)	MW_{SAXS} range (kDa)	credibility MW_{SAXS} (%)	MW_{MALS} (kDa)
PRC	76656.4 ± 69.7	10.0 ± 0.6	32.5	3320	829.5 - 1,237.5	92.0	880.0 ± 0.1%

The results for MW estimations correlate with previous observations made in this work (compare chapter GraFix and Blue native-PAGE 4.4.3, *in situ* DLS 4.4.4 and SEC-MALS 4.4.5). However, exclusive MW-estimations are not suitable to prove a nonameric or decameric arrangement of the subunits here. The distance distribution is almost bell shaped but shows an overlap of two bell shaped curves. This indicates that the proline reductase particle has an overall globular shape featured with cavities and smaller as well as globular elements. The Guinier range was set to cover the data points 10 to 40 to cover the lowest scattering angle possible and not to exceed an sRg of 1.3. A set of 20 3D bead models was generated, predefining all possible and rational symmetries. The results were compared and assumption of a 5-fold symmetry led to the most credible results with a Chi^2 of 1.183. A 5-fold symmetry, as it is shown in figure 4.26, supports a decamer of PrdAB-hetero trimers. This would consist of five hexamers that have a 2-fold symmetry themselves, as it had been found in the chapters 4.3.2 and 4.3.3. Therefore, the SAXS-results suggest that the proline reductase complex is a triacontamer. The models were used for later comparisons to negative stain TEM projections.

4.4.7 Negative stain TEM with the proline reductase complex

Due to the large size of the proline reductase complex, negative stain TEM as well as cryoEM are suitable techniques for structural investigations. Although high-resolution data of any proline reductase subunits are not yet available, our understanding of the exact stoichiometry and particle composition improved significantly by summarizing results from MALS, SAXS, *in situ* DLS and diverse crosslinking experiments. Also, complex stability

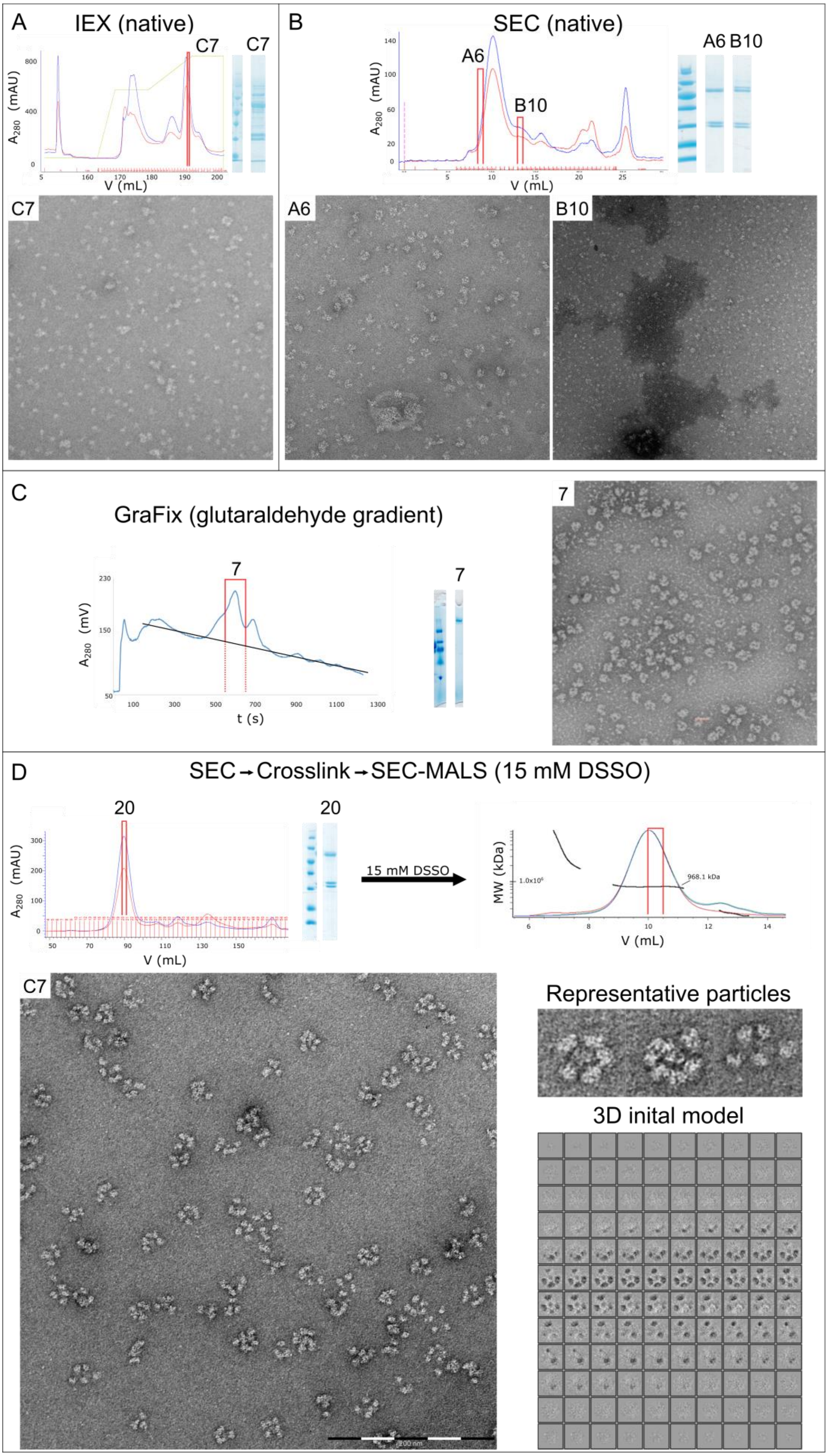

A
IEX (native)
C7
C7
A_{280} (mAU)
800
400
0
5
160
170
180
190
200
V (mL)
B
SEC (native)
A6
B10
A6 B10
A_{280} (mAU)
140
100
60
20
0
5
10
15
20
25
V (mL)
C7
A6
B10
C
GraFix (glutaraldehyde gradient)
7
7
A_{280} (mV)
230
150
50
100
300
500
700
900
1100
1300
t (s)
7
D
SEC→Crosslink→SEC-MALS (15 mM DSSO)
20
20
A_{280} (mAU)
300
200
100
0
50
70
90
110
130
150
V (mL)
15 mM DSSO
MW (kDa)
1.0x10^6
968.1 kDa
6
8
10
12
14
V (mL)
C7
200 nm
Representative particles
3D inital model

Figure 4.27: Comparative negative stain TEM data on the proline reductase supercomplex. A) IEX purified proline reductase is collected and plotted for negative stain TEM. The micrograph shows inhomogeneity and disintegration. B) A subsequent SEC isolation improves purity, as indicated by SDS-PAGE. However, the micrograph quality is not improved. C) Proline reductase is applied to GraFix, where the sample is crosslinked by GA, as it migrates into the glycerol gradient. The resulting micrograph shows improved size distribution of the particles. However, no clear preferred conformations can be observed. D) Crosslinking of the SEC-purified proline reductase with 15 mM DSSO and subsequent SEC-MALS analysis leads to micrographs, that are free of protein fragments. Also reoccurring patterns can be observed. Certain projections show a clear 5-fold-symmetry. However, 2D-classification does not lead to evolution of higher resolution details. A first ab initio 3D model shows common patterns but is not refinable. All micrographs are prepared with 4% UAc staining. Images are collected at HZI/ZEIM on a Carl Zeiss Libra 120 electron microscope.

and reasons for protein disintegration became more transparent, so that treatment during grid preparation could iteratively be adjusted in parallel to ongoing experiments.

The negative stain TEM grid preparation was performed by Mathias Müsken, HZI Braunschweig, Central Facility for Microscopy. Samples to be blotted on TEM grids were taken after preparative IEX, GraFix (GA supplemented density gradient centrifugation), SEC and after crosslink-SEC-MALS. The grid quality improved iteratively and final experiments allowed single particle picking, as one can see in figure 4.27.

Non-crosslinked proline reductase did not lead to micrographs of sufficient quality. Figure 4.27 A) and B) prove sample heterogeneity and disintegration. Fixation of the complex by GraFix (see figure 4.27 C) lead to micrographs where the size distribution of the particles seems to be improved. However, the images appear blurry and the particles do not seem to establish homogeneous conformations. Best results were achieved after crosslinking the proline reductase supercomplex with 15 mM DSSO for 30 minutes at RT and subsequent SEC-MALS isolation, as shown in figure 4.27 D). Eluting material was fractionated and subjected to negative staining immediately after collection. The micrographs prove the proline reductase supercomplex to be present as a high MW complex with more than 20 nm in size. This observation is in line with previous experiments, as discussed in the chapters 4.4.4 to 4.4.6. Moreover, it became obvious for certain projections that the proline reductase supercomplex assembles with a 5-fold symmetry, which is in agreement with SAXS-data as well (see chapter 4.4.6). However, particle homogeneity was considered as low, as 2D classes appeared diffuse and *ab-initio* 3D reconstructions showed very low resolution. Possibly, a flexible linker region, connecting the central oligomerization domain with the five catalytic domains is responsible for the poor model quality.

4.4.8 Coevolution Analysis with the proline reductase complex

Performed experiments showed that the proline reductase complex is assembled as a triacontamer. The three proteins, PrdAα, PrdAβ and PrdB form a hexamer that is referred to as subcomplex. The supercomplex is assembled from five subcomplexes, if the N-terminus of PrdAβ is present. To substantiate the observations made, sequence analysis was performed, including a HHBlits sequence alignment (Remmert et al., 2012) and subsequent GREMLIN analysis for the identification of coevolution clusters (Ovchinnikov et al., 2014). Coevolution occurs in protein-protein-interfaces that would lose their interaction ability if only one participating amino acid would undergo mutation. In order to maintain the functionality of the interface, the interacting amino acid eventually adopts with a second mutation, which is referred to as coevolution.

Conservation of domains was illustrated using Jalview (Clamp et al., 2004). To make sequences of the proline reductase complex accessible for GREMLIN-analysis, it needed to be divided in sections, as the number of related and homologous proteins for the proline reductase is rather small. First results were generated for PrdA 1 - 200. Jackhmmer search aligned 120 sequences, from which 12 were removed due to a coverage below 75%. Therefore, an overall seq/len value of 0.6 was achieved, leading to potentially noisy data. For this reason, an analysis covering the whole length was not possible. In a first attempt, the first 200 amino acids were subjected for analysis, as shown in figure 4.28. The graph revealed that this sequence stretch holds three clusters with agglomerations of coevoluted residues. The N-terminal “Cluster I” for the residues 1 - 100 was accompanied by a high degree of sequence conservation. A further mentionable region was found in the range from amino acid position 110 to 125, showing coevolution to residues spanning a large stretch of the N-terminus. According to secondary structure prediction (Jones, 1999; Buchan et al., 2019), this stretch describes a 16 amino acid long β-strand, which possibly includes the whole N-terminal domain. The next “Cluster II” of coevoluted residues was found to range from amino acid position 170 to 200. This sequence stretch is already present in the truncated version of PrdA and presumably depicts the dimer-interface.

For more reliable results and deeper analysis, the first 100 amino acids, which presumably depicting the decamer interface, were reused for another GREMLIN analysis (compare figure 4.29). Further, the PrdA sequence was truncated at the elastase cleavage site A149 and the serinolysis cleavage site T420 for closer investigations. Here, the sequence of PrdB was added to the search for inter-protein interactions. Results are shown in figure 4.30.

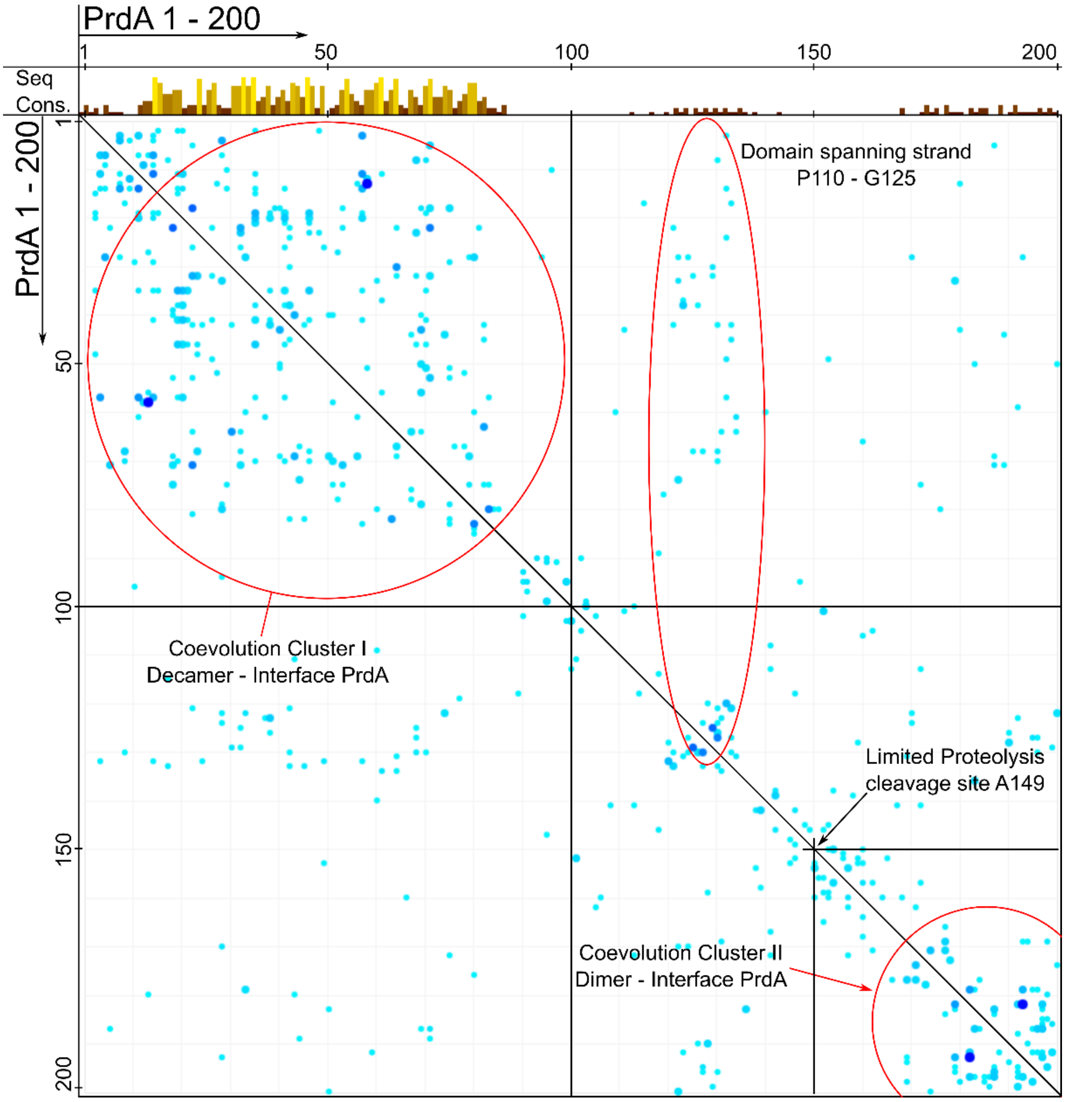


Figure 4.28: GREMLIN analysis for conservation and coevolution in PrdA$_{1-200}$. The darker and larger the blue dots are, the higher the degree in coevolution is. The plot displays residues in the sequence that show coevolution based on a BLAST search. This analysis allows predictions of interaction sites, even if the 3-dimensional structure is unknown. For PrdA, the stretch of the first 100 amino acids shows a dense cluster of interacting amino acid positions. This area is also well conserved among homologs. Another interesting region is found from amino acid position 110 to 125, which is predicted as a long β-strand and shows coevolution to amino acids covering the first 100 N-terminal residues. Further, a sequence stretch starting with amino acid 160 displays another cluster of coevoluted residues.

The resubmission of the PrdA sequence 1 to 100 (see figure 4.29), covers a conserved domain with a high amount of coevoluted residues. In more detail, 18 pairs of amino acids were identified with a calculated coevolution probability of more than 75%. The amino acid pairs are given in table 4.9. The properties of involved residues include small and bulky amino acids, as well as acidic, basic, hydrophobic and aromatic residues. Despite the pair H11-D14 found with a probability of 96%, no potential ionic interactions have been found. The residues identified are predominantly hydrophobic, which are typically located in the hydrophobic core of a protein or in chaperone generated interfaces.

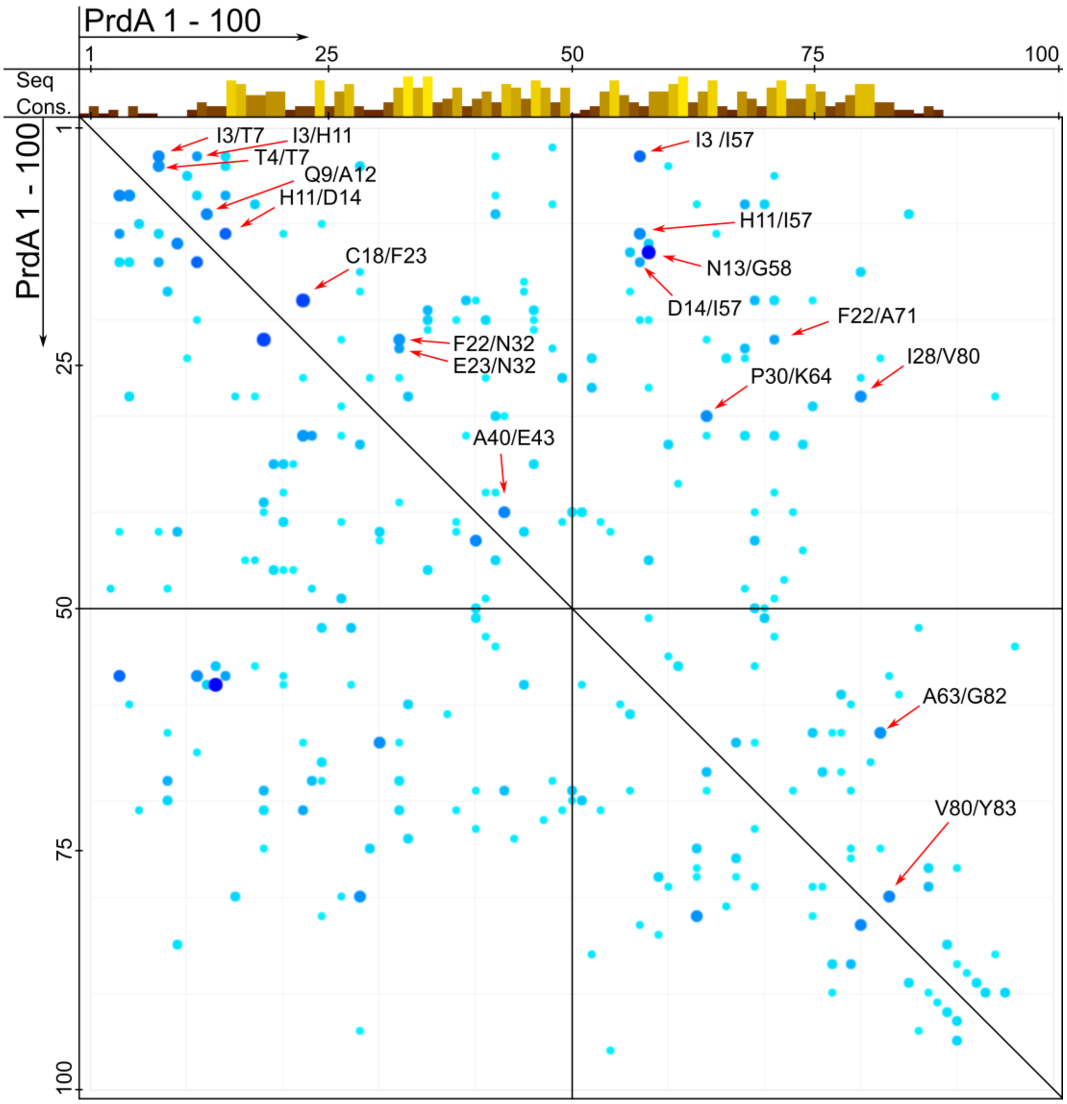


Figure 4.29: Detailed GREMLIN analysis of coevolution cluster I. The cluster I is located within the first 100 amino acids of PrdA. Blue spot size and intensity represents the coevolution strength. This area shows a high number of residues that potentially interact, as visualized by red arrows. The prediction is based on a sequence comparison, with an E-value of 10^{-4}, with 8 iterations and 50% coverage.

The results for the GREMLIN analysis on the PrdA sequence ranging from amino acid position 1 to 100 illustrated in figure 4.29 are summarized in table 4.9.

Table 4.9: Summary GREMLIN coevolution analysis for PrdA. Coevoluted residues in the range of amino acid position 1 to 100. E-value 10^{-4}, with 8 iterations and 50% coverage are shown.

Pos1	Pos2	Probability	Pos1	Pos2	Probability
N 13	G 58	1.000	P 30	K 64	0.854
C 18	F 22	0.988	A 63	G 82	0.843
H 11	D 14	0.961	I 28	V 80	0.839
I 3	I 57	0.957	T 4	T 7	0.834
V 80	Y 83	0.885	F 22	N 32	0.821
A 40	E 43	0.883	I 3	H 11	0.812
Q 9	A 12	0.880	E 23	N 32	0.793
I 3	T 7	0.875	D 14	I 57	0.762
H 11	I 57	0.865	F 22	A 71	0.753

Based on the conservation pattern and the density of coevoulting residues, the amino acid stretch M1 to Y83 is considered to depict the decamerization domain. Interestingly, this amino acid stretch is accompanied by a low net charge due to a high glutamate content (EMBOSS analysis, see figure 4.4) and is immediately followed by a flexible region (DISOPRED, see figure 4.2), which strengthens this hypothesis.

The generation of P2-symmetrical subcomplexes was thought to be mediated by an amino acid stretch beginning at position A149 of the PrdA sequence (compare chapter 4.1). PrdA$_{149-626}$ was further found to interact with PrdB$_{U151C}$. Possible interfaces were again addressed by sequence submission to the GREMLIN server, as illustrated in figure 4.30. Due to low number of homologs, the sequence of PrdA needed to be C-terminally truncated to position T420.

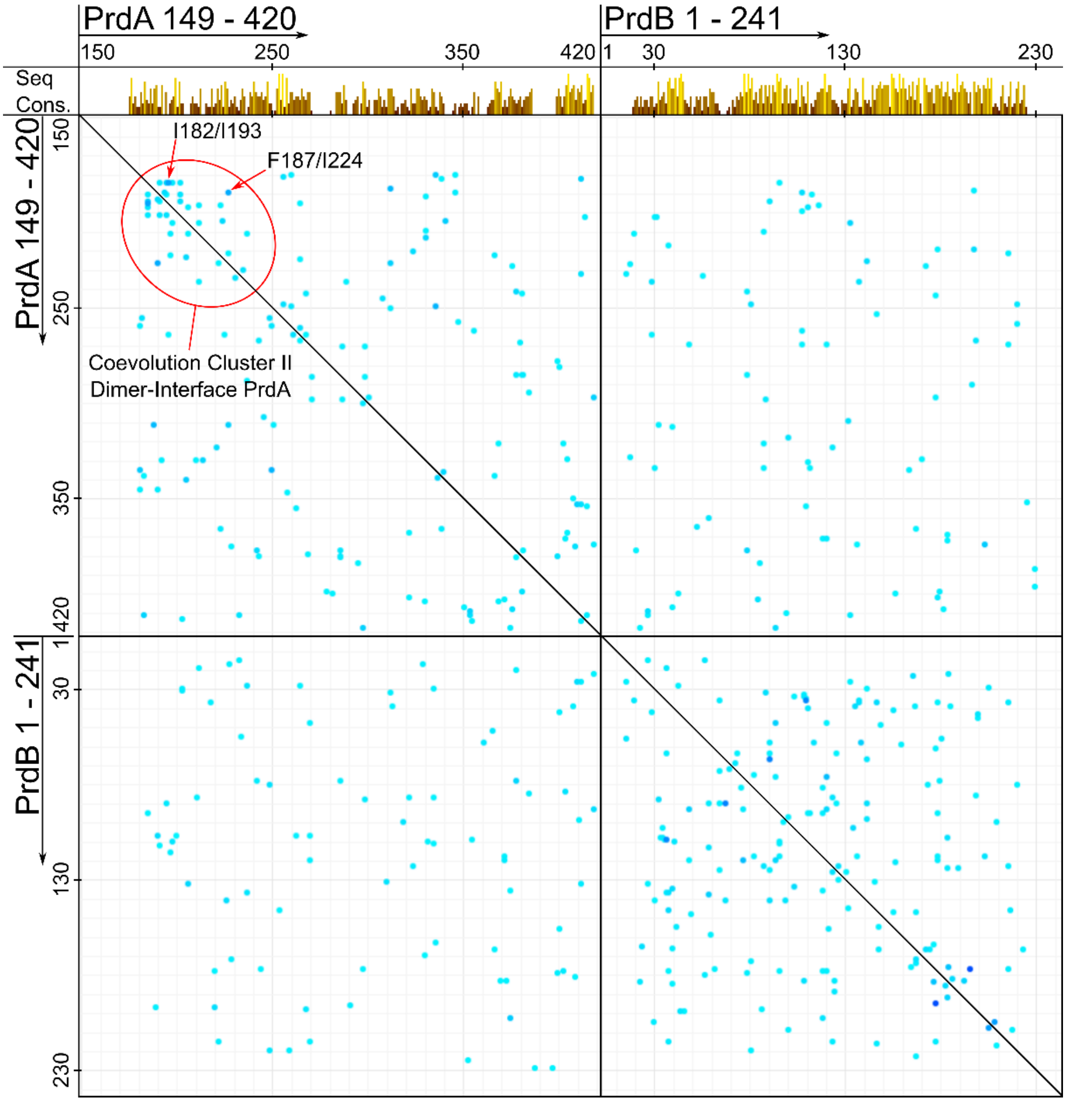


Figure 4.30: Detailed GREMLIN analysis of coevolution cluster II. This cluster is found within amino acid positions N150 to T420 of PrdA and PrdB. Blue spot size and intensity represents the coevolution strength. The second cluster of coevolution is smaller than cluster I and is approximately restricted to the amino acid positions from N150 to E250. No coevolution was between PrdA N150 – T420 and PrdB was observed that would give detailed information on the interface. For PrdB only, some residues showing coevulution were identified that must be caused by intramolecular interactions. The prediction is based on a sequence comparison, with an E-value of 10^{-4}, with 8 iterations and 50% coverage.

GREMLIN sequence analysis of PrdA identified an agglomeration of coevoluted residues, referred to as “Cluster II”. However, in comparison to Cluster I, the sequence conservation is lower. Hence, prediction accuracy suffers and only two pairs of coevoluted residues show a probability above 50%, which are the pairs I181/I193 and F187/I224. Further, no significant coevolution between $\mathrm{PrdA}_{149\text{-}420}$ and PrdB was detected. As a consequence, PrdA/PrdB-interaction is herewith predicted to be realized in the PrdAα subunit. In analogy to Cluster I, cooccurrence of flexible regions (compare figure 4.2) and glutamate rich region (compare figure 4.4) was observed, again confirming previously suggestions that glutamate bound ions are needed for complex integrity.

4.5 Summary - proline reductase complex

Putting together observations made on full length PrdA (see chapter 4.1), certain peculiarities become clearer that partially explain occurring difficulties. This is true for sample handling and poor behavior of both, PrdA and the proline reductase supercomplex. Some sequence stretches in PrdA can now be classified as domains with different functions. Removing the N-terminal domain ranging roughly from AA M1 to V148, as it was found by limited proteolysis (compare chapter 4.1.3), the remaining AA A149 – N626 are still able to form a homo dimer of $\text{PrdA}_{149\text{-}626}$, exhibiting a C_2-symmetry. However, proline reductase supercomplex formation is prevented and previously observed agglomeration was diminished. Therefore, this N-terminal domain can be referred as a decamerization domain. It is distinguishable from the remaining PrdA by a high content of glutamines in the range of AA 1 to 100 and therefore shows strong local negative charges due to deprotonation, as visualized by EMBOSS (compare figure 4.4). Indications for ion-mediated oligomerization were found in a TSA experiment, where tremendous destabilization of $\text{PrdA}_{149\text{-}626}$ by chelating agents like EDTA, EGTA and TCEP was observed. Again, glutamines can be found in protein-protein interfaces and might be essential for ion-binding and oligomer formation (Krebs, 1967; Stott et al., 1995; Sharp, 1996; Jeong et al., 2002; Pace et al., 2014; Garapati et al., 2018). The transition from negative to neutral charge is accompanied by a confidence of flexibility and protein binding prediction in the range of AA V80 – V120 (compare figure 4.2). Subsequent sequential analysis of this N-terminal stretch confirms the hypothesis of an oligomerization domain by identification of coevoluted residues (compare figures 4.28-30). This complicated architecture likely causes poor sample quality for recombinant PrdA production. Possible solutions could be the transfer to a Gram-positive expression system, such as *Bacillus megaterium* or coexpression from the whole *prd*-operon. Such an approach has been shown to be valid for protein complex production and in some cases even lead to subsequent structure elucidation (Selleck et al., 2008; Poulsen et al., 2010; Babnigg et al., 2015).

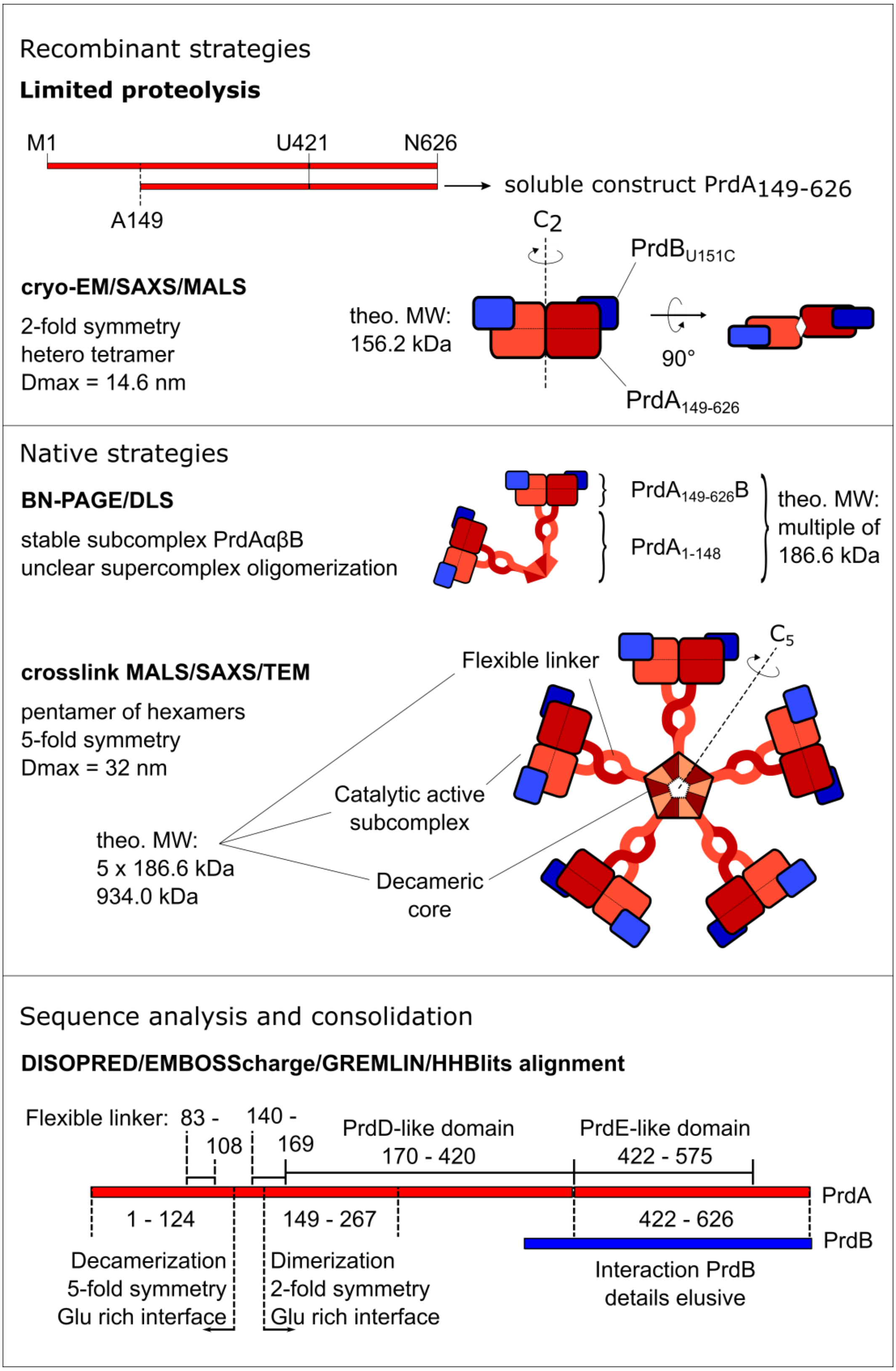


Figure 4.31: Conclusions proline reductase supercomplex. All experiments performed draw a coherent image of the *C. difficile* proline reductase supercomplex, which is summarized here. Limited proteolysis circumvented formation of soluble PrdA aggregates. The interaction to PrdB$_{U151C}$ was maintained.

The second domain of PrdA is referred to as PrdD-like domain, due to its sequence similarity to PrdD from AA E167 to T420 with 41.1% identity. PrdD is another protein encoded in the *prd*-operon with yet unknown function. The PrdD-like domain starts with the first positively charged section of PrdA from AA A149 - F180, followed by a second, large negatively charged region, again promoted by a high content of glutamines in the range of AA F180 – E250. Putatively, this stretch of PrdA is involved in dimer formation. Combining experimental results and sequence analysis, the dimerization domain is thought to cover the amino acid region from position A149 to T267. The remaining stretch ranging from G268 to T420 cannot be annotated to a function yet. Present data does not suggest a role in oligomerization or protein-protein-interaction. One may only hypothesize a role in proteolytic maturation or enzymatic activity, as this domain concludes with the serinolysis cleavage site.

The following domain in PrdA (I422 to V575) will be referred to as PrdE-like domain, as it shares 44.5% identity and 67.1% similarity with PrdE and covers more than 150 amino acids. The PrdE-like domain carries the N-terminal pyruvyl group in mature PrdA. This domain holds the only sequence motif that can be used for sequence based structural prediction by Phyre2. The two most promising results depict asparaginases, that adopt an α/β-fold (Lubkowski et al., 1996; Sanches et al., 2003). Interestingly, those hits turn over amino acids, which was also found to depict a characteristic property of pyruvoyl-dependent enzymes. Both properties are true for the proline reductase (Ekstrom et al., 2001; Schmitzberger et al., 2003; Tolbert et al., 2003; Andreesen, 2004; Bale et al., 2010; Trip et al., 2011; Jiang et al., 2013). Further this domain is thought to hold the PrdB-interaction-site, since none of the other domains investigated showed indications for an interaction with PrdB.

For PrdB itself, a concrete basis for structural elucidation was created. Still, despite partially promising results in gained sample amount and quality, reproducibility in batch generation could not be achieved. Therefore, no data suitable for structural elucidation of PrdB could be collected by X-ray crystallography or NMR spectroscopy. Most likely, obstruction was caused by an N-terminal amino acid stretch with high flexibility and multiple structural conformations. It was found by combination of CD spectroscopy, tryptophane fluorescence emission shift spectroscopy and NMR spectroscopy that PrdB adopts at least two distinguishable conformations, which are only insufficiently separable by SEC. The biological relevance of those species remains elusive. The problem must be addressed not only by means of purification, but probably also by testing various expression conditions. Tryptophane fluorescence emission shift spectroscopy provides a rapid and robust method to validate folding states of recombinant PrdB. Eliminating the source of this complication is essential for further studies and structural elucidation of PrdB. Most promising attempts are given with NMR spectroscopy, which was proven to work in principle on PrdB (see figure 4.7). Once the sample shows homogeneity and chemical shifts could be annotated, the interaction sites to PrdA can be easily determined.

One milestone in *C. difficile's* proline reductase related research was passed by validating cryo-EM as a method suitable for data generation on the subcomplex of truncated PrdA and PrdB. This complex could be produced recombinantly in high amounts and good purity. It has been proven to be stable and was monodisperse in solution. Characterization of the complex was addressed by SEC-MALS, SEC-SAXS and cryo-EM, drawing a consistent image of a rather rigid heterotetramer, which exhibits a C_2-symmetry with 154 kDa. Future experiments will include repetition of cryo-EM data acquisition at different specimen tilting angles and addition of detergent or other compounds to native and crosslinked $PrdA_{149-626}PrdB_{U151C}$ to overcome preferred particle orientation as observed in chapter 4.3.3 (Scheres, 2016; Zi Tan et al., 2017; Drulyte et al., 2018; Chen et al., 2019). This approach will hopefully lead to isotropic resolution of the subcomplex by cryo-EM single particle analysis. Potentially the map could be reconstructed to a resolution of more than 4 Å, as it had been observed to be true for the projection along the C_2-symmetry axis. Additionally, further truncation versions of PrdA are under investigation to promote crystallization.

Natively purified proline reductase complex has been accessible before and purification protocols were refined in order to improve sample homogeneity and stability (Seto et al., 1976; Kabisch et al., 1999). The PrdAαβB-supercomplex was known to be present as a high MW assembly, lately described as a decamer or nonamer of trimers. However, structural assembly and details were elusive and had not been addressed for a longer period. With occurrence of new process resins, exhibiting improved binding properties, revisiting the formerly used HIC material has been a promising approach. Indeed, testing various resins lead to identification of a Butyl-HP resin, which had the highest affinity for the mature proline reductase complex among all resins tested, when 1 M $(NH_4)_2SO_4$ was present in the buffer. As a consequence, precipitation of the complex became obsolete. The obtained protein mixture could easily be purified by MonoQ IEX, resulting in an almost pure sample after pH and salt gradient optimization. The remaining impurity, the elongation factor G FusA, could effectively be removed by SEC, using Superose 6 material. Combination of different bioanalytical approaches lead to the conclusion that the subcomplex of PrdAαβPrdB shows a high stability. However, the supercomplex assembly depicts a flexible and dynamic process, which is not controllable by optimization of purification procedures or buffer composition. Instead, reliable stabilization was achieved by crosslinking the supercomplex with 15 mM DSSO, an amine-reactive, monofunctional crosslinker with a linker distance of approximately 10 Å. The stabilized sample could finally be addressed by SEC-SAXS, SEC-MALS and negative stain TEM, suggesting that the supercomplex is a pentamer of hexamers. Structural investigations on the supercomplex by native cryo-EM were not successful, as the complex showed heavy disintegration. Low resolution negative stain TEM was optimized and the C5-symmetry was visually recorded. Application of the optimized crosslinking procedure for cryo-EM grid preparation could potentially lead to successful data collection and processing. However, due to observed flexibility it is likely that resulting maps will suffer from poor resolution. Still, even a low-resolution map will be helpful for map fitting of the $\mathrm{PrdA}_{149\text{-}626}\mathrm{B}_{\mathrm{U151C}}$ subcomplex to correctly localize enzymatically active subcomplexes in the context of the supercomplex structure. Urgent questions are the orientation of catalytic centers within the subcomplex and possible effects on reactivity, as well as the exact orientation of the catalytically active domains towards each other in reduced and oxidized forms. Simultaneously, the decameric core complex, presumably mirroring the C_5-symmetry, will be defined as a target for structural investigations. Structural information of the decamerization domain could potentially shed light on the relationship of oligomerization and biological function of the proline reductase complex.

5 Topic II: Structure of the Phospholipase PlaA from *Legionella pneumophila*

Purified Strep-PlaA was provided by MSc Miriam Hiller from the Robert Koch Institute, Wernigerode. The material was subjected to a SEC polishing step and crystallization. Crystallization optimization, co-crystallization and soaking experiments were conducted to elucidate the determined structure of PlaA.

5.1 Crystal structure of PlaA

The purified protein was polished by gel filtration using a S200 10/30 *Increase* column, pre-equilibrated with PlaA SEC buffer. After SDS-PAGE, the fractions containing the main peak were pooled and concentrated to 8 mg/mL before subjection to crystallization, as shown in figure 5.1.

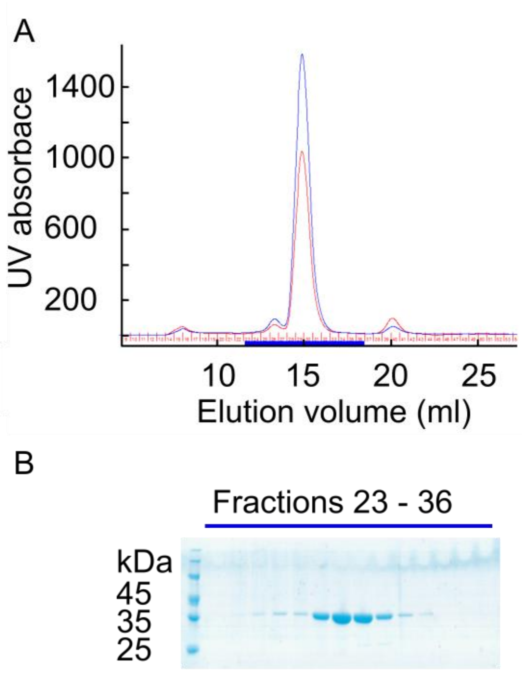


Figure 5.1: SEC polishing of PlaA. A) The provided protein was purified by size exclusion chromatography, using a S200 10/30 Increase column. The fractions marked with a blue bar were collected and analyzed by SDS-PAGE. B) The SDS-PAGE of collected fractions shows one single band corresponding to PlaA. Fractions containing PlaA were pooled and concentrated to 8 mg/mL for crystallization experiments.

Initial screening using a JCSG+ screen led to small crystals from several conditions containing PEG in different sizes. Therefore, optimization was performed with the aim to enhance crystal size. However, no high-quality crystals were achieved. It was hypothesized that crystallization in initial screens was successful due to the presence of smaller PEG degradation products. Therefore, further optimization was performed by mimicking PEG degradation. This was achieved by addition of small PEG's, i.e., ethylene glycol, PEG200, PEG 1K and Tacsimate, a mixture of organic acids adjusted to different pH-values. This approach was successful and crystallization behavior improved. Interestingly, the addition of the PlaA substrate lysophosphocholine (LPC) (16:0) made the presence of smaller PEG's and Tacsimate unnecessary for crystallization.

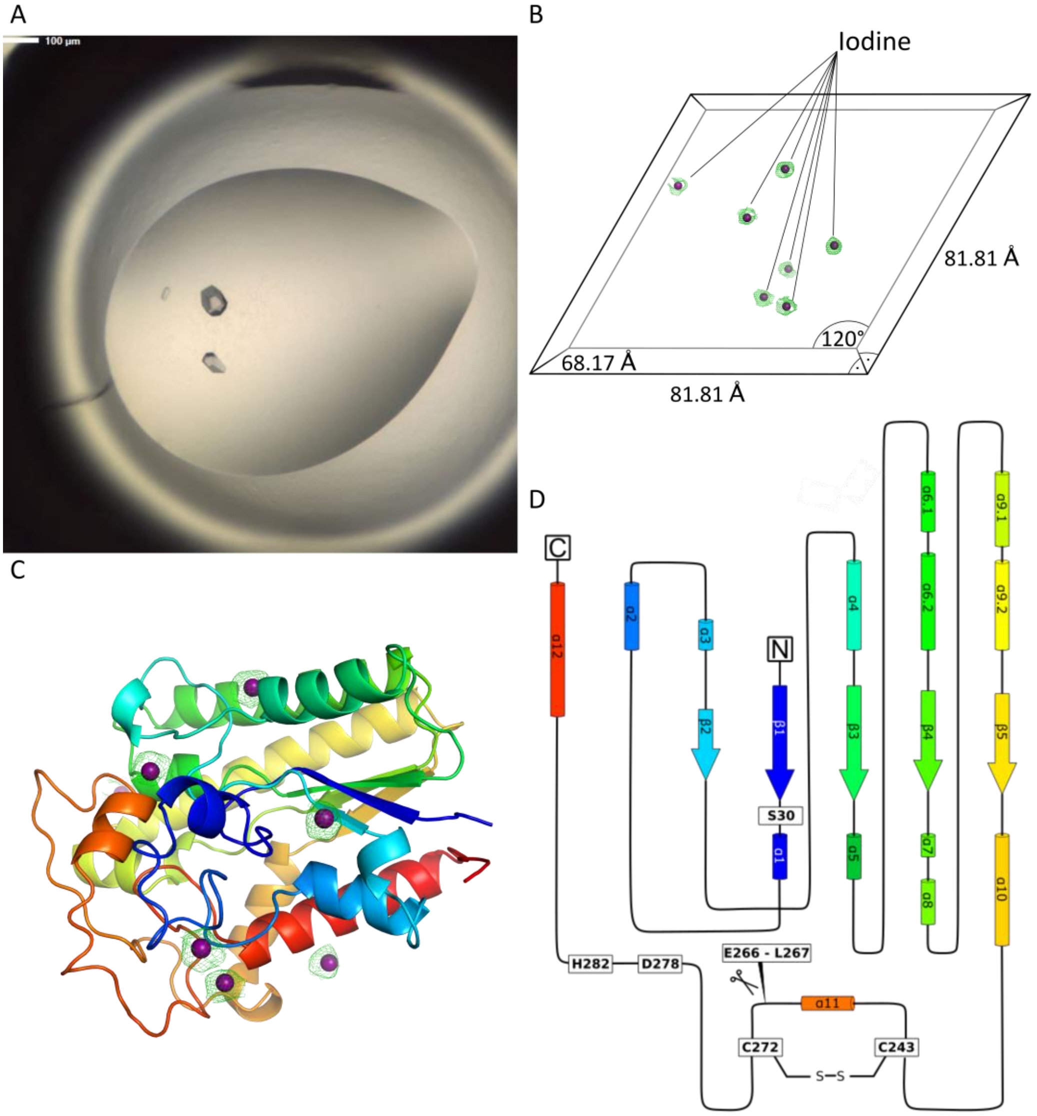


Figure 5.2: Structure determination of PlaA. A) Crystals grew in a condition containing 14.7% (w/v) PEG3350, 1% (w/v) PEG 1K, 0.5% (w/v) PEG200, 8.9% (v/v) Tacsimate pH 5 and 200 mM NH_4I, 1 mM $CaCl_2$ and 1 mM $MgCl_2$. B) After phase determination, seven iodine atoms were placed in the ASU with the parameters 81.81/81.81/68.17/90/90/90. The green electron density shows the anomalous difference map at σ = 1. C) One molecule of PlaA was placed in the ASU. D) The topology diagram shows the arrangement of SS elements as a 2-dimensional projection. The active site and termini are highlighted in white boxes. The fivefold central β-sheet is surrounded by a set of α-helices. The lid is defined as the stretch between a cysteine bridge ranging from C243 to C272 and is present in a closed conformation. Helix α11 depicts the prominent element of the lid, covering the active site effectively.

For iodine phasing, PlaA crystals were grown in 14.7% (w/v) PEG3350, 1% (w/v) PEG 1K, 0.5% (w/v) PEG200, 8.9% (v/v) Tacsimate pH 5 and 200 mM NH_4I, 1 mM $CaCl_2$ and 1 mM $MgCl_2$. The crystal was fished and soaked in a solution of 10% w/w 1,2-butandiol, 16% (w/v) PEG3350, 6% (v/v) Tacsimate pH 5 and 920 mM NH_4I. Two datasets with a detector shift (distance 140 mM and 135 mM) were collected at the SLS, Villigen, Switzerland (Bingel-Erlenmeyer et al., 2011). The energy was set to the lowest energy possible of 7.000 keV (corresponding to wavelength λ = 1.771196 Å), exploiting the f'/f'' difference arising from the L-I-absorption edge of iodine (5.1881 keV, 2.3898 Å). Both DS were collected with an oscillation range of 0.1° and 360° in total. The DS were processed on site by XDSGO and merged with XSCALE (Kabsch, 2010 a). Phases were determined with a strong anomalous signal up to 2.50 Å resolution and an overall 20-fold multiplicity by hkl2map and the SHELX-package (Pape et al., 2004). The FOM was determined to be 0.779, indicating a successful phasing procedure. Subsequently, an initial poly-alanine model was generated, which was used later as a reference model. Seven iodine atoms in total could be positioned in the map.

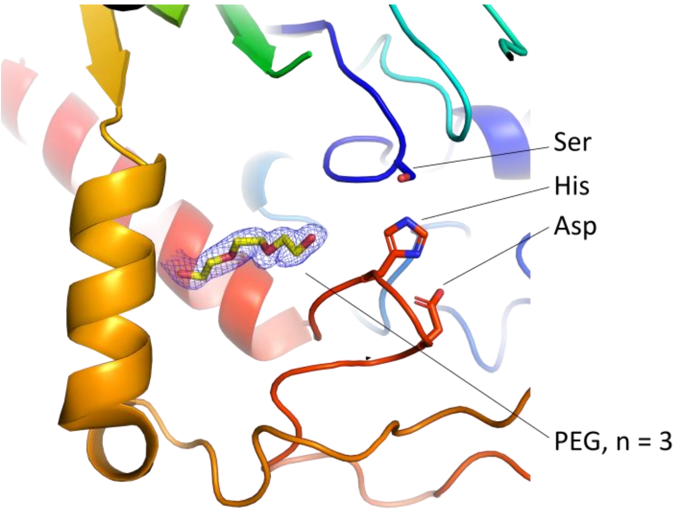


Figure 5.3: Ligand coordination in the active site of PlaA. After refinement, additional density was found in the center of PlaA, filling the space in a hydrophobic pocket in proximity to the active site. Due to small PEGs in the crystallization condition, this density was considered to arise from PEG (n = 3)

The observed behavior in crystallization already suggested that smaller PEG molecules are essential for crystallization of PlaA. Indeed, after refining the model, additional electron density was found in the center of the protein, as shown in figure 5.3. Due to its elongated shape, it was interpreted to arise from short PEG molecules. As the coordinated PEG was located in a hydrophobic channel in close proximity to the active site, a mimicking mechanism was assumed. The PEG degradation product shows similarity to a long chain fatty acid, one of the reaction products of LPC hydrolysis.

Table 5.1.: Data collection and refinement statistics PlaA. Two datasets are collected from a PlaA crystal. Data collection was performed at the SLS at low energy to record f’-f’’-differences remote from Iodine L-I absorption edge. In between the datasets the detector distance was shifted from 140 mM to 135 mM. Datasets were merged by XSCALE and phases were determined by hkl2map. Refinement was performed using phenix.refine. Statistics for the highest-resolution shell are shown in parentheses.

		DS1 (140 mM)	**DS2 (135 mM)**	**Merged statistics**
Data collection	**Wavelength [λ]**	1.771	1.771	
	Resolution range [Å]	49.3 - 2.3 (2.4 - 2.3)	49.3 - 2.4 (2.5 – 2.4)	10.3 – 2.3 (2.4 – 2.3)
	Space group	P 3 (143)	P 3_1 2 1 (150)	P 3_1 2 1 (150)
	Unit cell a/b/c [Å]	81.8/81.8/68.2	81.9/81.9/68.2	81.8/81.8/68.2
	Unit cell α/β/γ [°]	90.0/90.0/120.0	90.0/90.0/120.0	90.0/90.0/120.0
	Total reflections	226470 (33141)	217967 (33848)	444207 (19394)
	Unique reflections	22580 (3631)	21455 (3469)	22580 (1691)
	Multiplicity	10.0 (9.1)	10.2 (9.8)	19.7 (11.5)
	Completeness [%]	100.00 (99.90)	100.00 (99.9)	100.00 (99.7)
	Mean I/sigma(I)	11.21 (1.15)	11.20 (1.19)	15.78 (1.17)
	R-meas	17.3 (157.8)	17.6 (159.7)	16.6 (179.0)
	CC1/2	99.8 (53.8)	99.7 (56.3)	99.9 (50.1)
Refinement	**Reflections refinement**			42349 (3515)
	Reflections R-free			2115 (173)
	R-work			0.21 (0.49)
	R-free			0.23 (0.50)
	Number non-hydrogen atoms			2455
	macromolecules			2287
	ligands			15
	solvent			153
	Protein residues			290
	RMS(bonds)			0.01
	RMS(angles)			1.19
	Ramachandran favored [%]			96.18
	Ramachandran allowed [%]			3.12
	Ramachandran outliers [%]			0.69
	Rotamer outliers (%)			5.24
	Clashscore			16.27
	Average B-factor			32.29
	macromolecules			31.77
	ligands			65.49
	solvent			36.78

The presence of additional density in the protein core and the necessity of PEG for PlaA crystallization suggested that ligand addition played a key role in successful crystallization.

As a substrate bound ligand structure was desired to unveil reaction mechanisms, the substrate LPC (16:0) was added to an inactive mutant of PlaA and crystallization was optimized. In this experiment, any PEG was avoided to prevent competitive insertion of PEG molecules in the active site. In detail, a high-resolution structure of PlaA was achieved by adding 500 µM LPC (16:0) to a 7 mg/mL PlaA solution. The crystallization condition was 2.2 M $(NH_4)_2SO_4$ and 0.2 M NH_4-acetate. The previously experimentally determined structure of PlaA served as a MR search model. The structure is shown in figure 5.4. Statistics are summarized in table 5.2.

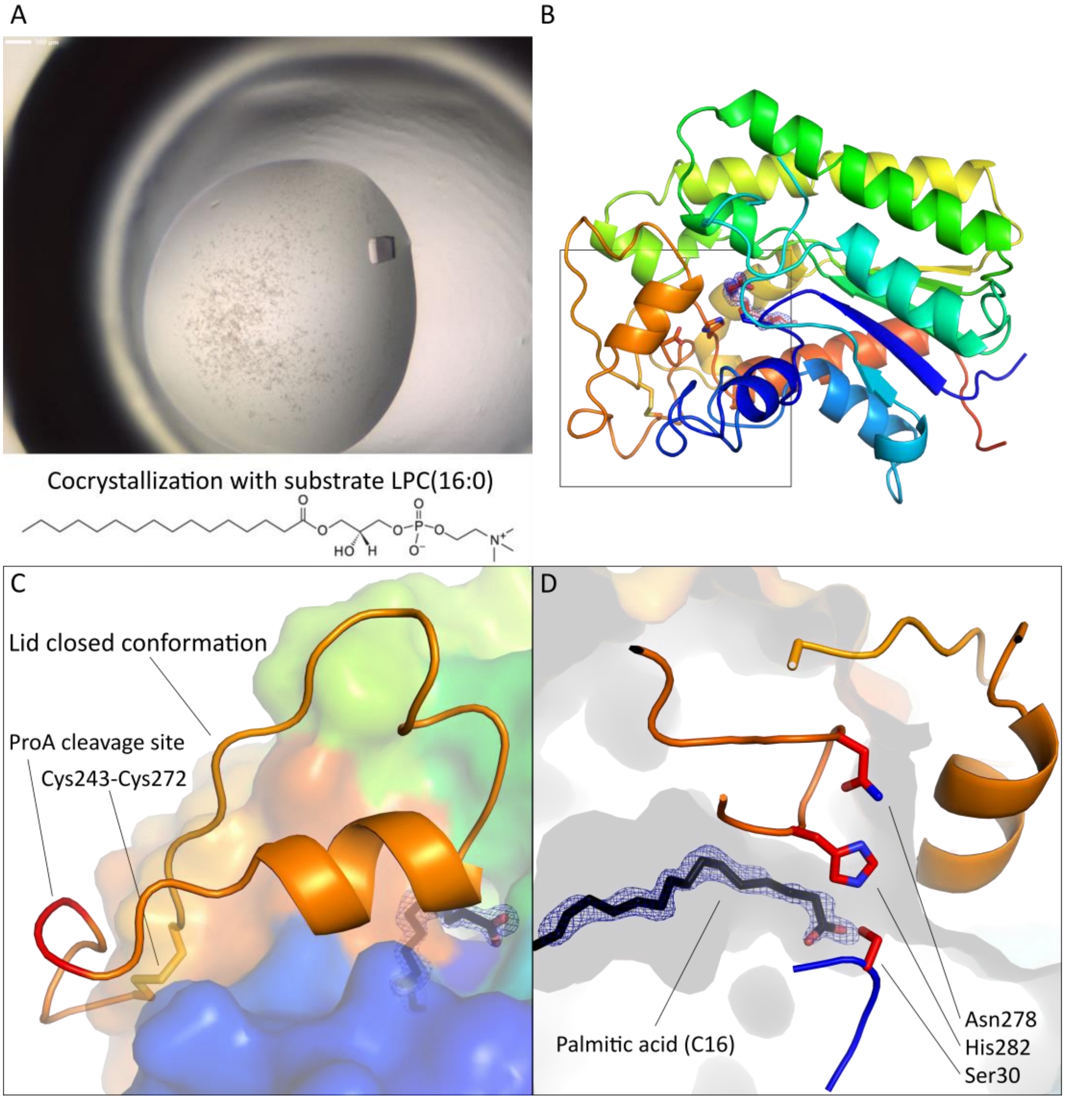


Figure 5.4: Crystal structure of PlaA in complex with palmitic acid. A) A single crystal grew upon addition of 500 µM LPC (16:0) from a 7 mg/mL protein solution in a crystallization condition containing 2.2 M $(NH_4)_2SO_4$ and 0.2 M NH_4-acetate. Data was collected at DESY P11 and phases were determined by MR with a previously generated model. B) The high-resolution structure of PlaA shows the same structure compared to the experimentally determined model (RMSD = 0.122 over 1817 atoms). C) The lid of PlaA was found in a closed conformation. The helix α11 efficiently shields the substrate tunnel from the solvent. The ProA cleavage site for maturation of PlaA is depicted in red. The stabilizing cysteine bridge Cys243 - Cys272 is shown as yellow stick-model. D) Palmitic acid was found as a ligand bound to PlaA, occupying a hydrophobic channel. The carboxyl group was found in close proximity to the mutated active site Ser30, Asn278, His282.

Interestingly, despite of using the inactive mutant $\text{PlaA}_{\text{D278N}}$ for crystallization, the substrate LPC (16:0) reacted to palmitic acid and, presumably, glycerol-3-phosphocholine (G3P). This might be possible from residual activity or autohydrolysis of the compound. The substrate and G3P were not found to be present in the structure. However, the reaction product palmitic acid was clearly located in the hydrophobic binding tunnel. The carboxyl group was pointing towards the active site, indicating a correct and not a random insertion of the reaction product.

Table 5.2: Data collection and refinement statistics PlaA, ligand structure. A high-resolution dataset is collected from a PlaA crystal after co-crystallization with the substrate LPC (16:0). Data collection was performed at DESY P11 at a wavelength of 1 Å. Statistics for the highest-resolution shell are shown in parentheses.

Data collection		**Refinement**	
Wavelength [Å]	1.000	**Reflections used in refinement**	46400 (4574)
Resolution range [Å]	41.4 - 1.4 (1.5 - 1.4)	**Reflections used for R-free**	2386 (257)
Space group	P 3_1 2 1 (150)	**R-work**	0.17 (0.26)
Unit cell a/b/c [Å]	82.85/82.85/68.60	**R-free**	0.20 (0.27)
Unit cell α/β/γ [°]	90/90/120	**Number of non-hydrogen atoms**	2592
Total reflections	855868 (136795)	**macromolecules**	2314
Unique reflections	45174 (8501)	**ligands**	84
Multiplicity		**solvent**	194
Completeness (%)	99.99 (98.80)	**Protein residues**	290
Mean I/sigma(I)	29.4 (1.65)	**RMS (bonds)**	0.01
Wilson B-factor	16.58	**RMS (angles)**	1.28
R-meas (%)	6.1 (150.2)	**Ramachandran favored (%)**	97.20
CC1/2	100.0 (65.4)	**Ramachandran allowed (%)**	2.80
		Ramachandran outliers (%)	0.00
		Rotamer outliers (%)	5.98
		Clashscore	18.34
		Average B-factor	24.30
		macromolecules	22.76
		ligands	45.88
		solvent	33.27

PlaA is a canonical GDSL hydrolase with an α/β-hydrolase fold (Flieger et al., 2004), as confirmed by crystallization and experimental phasing, shown in figure 5.2. The core β-sheet consists of 5 β-strands and is surrounded by a set of α-helices. The structure unveiled a lid domain, ranging from C243 to C272. This lid was found in a closed conformation, as structural comparisons proved (Jeong et al., 2002; Secundo et al., 2004;

Schleberger et al., 2007; Khan et al., 2017). Interestingly, it is stabilized by a disulfide bond located at its boundaries. The proteolytic cleavage site is known to be present in an amino acid stretch at the C-terminal end of the lid (Lang et al., 2017). The access to the active site and the hydrophobic binding pocket is effectively shielded by α-helix α11. Hence, it appears likely, that proteolytic maturation and activation of PlaA by ProA induces high flexibility in the lid structure and enables access of substrates to the active site.

Considering the lid to be flexible upon maturation, substrate binding is not explained entirely yet. Insertion of the long hydrophobic C16 chain through a small whole is energetically unlikely. Hence, an opening mechanism is considered, that opens the hydrophobic binding tunnel laterally. This could easily be achieved by removing the C-terminal helix α12, as demonstrated in figure 5.5. Considering α12 to be flexible explains how the substrate binding event could be thermodynamically enabled.

Another feature of PlaA can now be described by the generated structure. If unmatured, PlaA shows an activity towards an alternative substrate, cholesteryl palmitate. While the biological relevance remains unclear, potential binding motifs on the surface of PlaA could be identified as shown in figure 5.6. However, co-crystallization with this substrate failed, as those hydrophobic surface residues are involved in crystal contact formation.

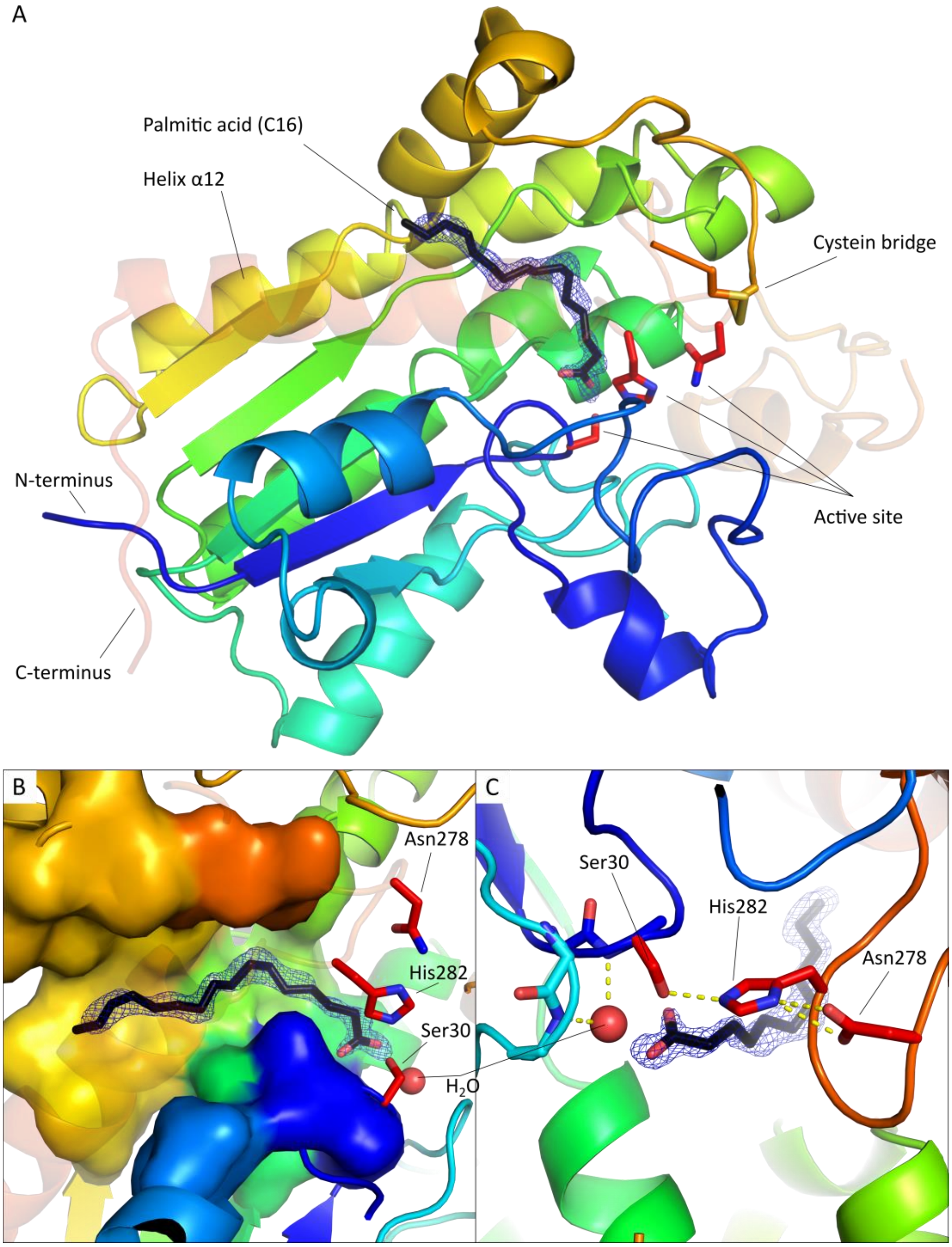


Figure 5.5: Ligand coordination in PlaA. A) Considering the C-terminus and α12 to be flexible before ligand binding, the hydrophobic binding tunnel becomes solvent accessible. This flexibility appears to be necessary for ligand insertion. B) The bound product appears in close proximity to the active site. However, in case of opening the entry site, the residues Asn278 and His282 are considered to be present in a disordered state as well. C) A water molecule is tightly coordinated by the backbone nitrogen of Ser30 and Gly86. The catalytic triad describes the proton subtraction by Asn278 from His282. The generated electron delocalization induces electron transfer from His282 to Ser30 and hydrolysis of the substrate. The water molecule is most likely to be bound in the oxyanion hole after successful subtraction of the electrons.

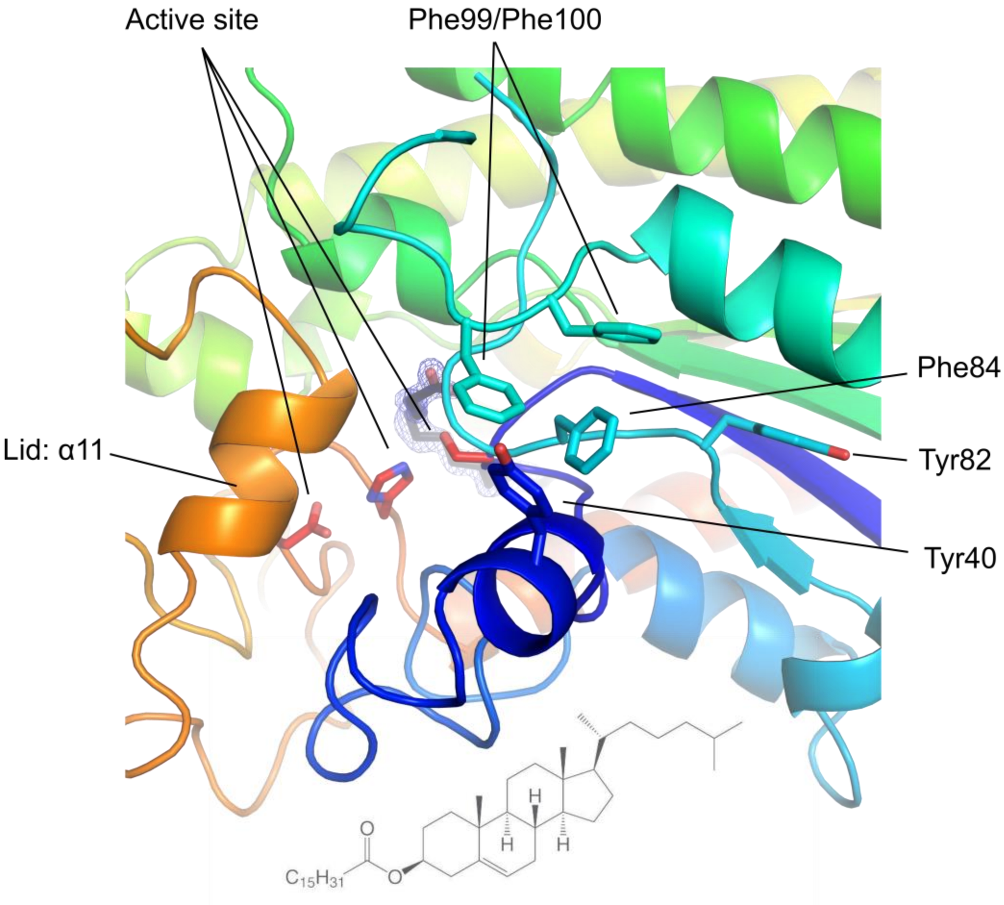


Figure 5.6: Atomic details of putative cholesteryl palmitate binding site. The aromatic residues Tyr40, Tyr82, Phe84, Phe99 and Phe100 are located on the surface of PlaA and depict potential binding sites for the ring system typically found in cholesteryl derivatives. The side chains are located in close proximity to the lid domain.

Assuming increased flexibility of the lid domain upon cleavage at position 266, a structural rearrangement depicts the most likely reason for activity induction towards the natural substrate LPC (16:0). Simultaneously, the same structural rearrangement could potentially diminish cholesteryl palmitate activity by sterical blocking of the surface located aromatic sidechains.

Together our data describe the relation of proteolytic maturation, activation and substrate specificity of PlaA. Upon secretion by the type II secretion system (T2SS), the metallo protease ProA (Hales et al., 1999; Truchan et al., 2017) cleaves the PlaA backbone in the range of the amino acids S260-L270, preferably at E266-L267 (Lang et al., 2017). The cleavage presumably brings a high degree of flexibility to the lid, which is structurally defined to cover the amino acid range surrounded by the disulfide bond C243 to C273. At the same time, the disulfide bond stabilizes the residual structural integrity of PlaA.

6 Topic II: Structure of the Phospholipase PlaB from *Legionella pneumophila*

To decipher the molecular basis for PlaB's unusual activation mechanism (compare chapter 1.2.4) and to gain insight into determinants of membrane association, the crystal structure of PlaB has been determined and corroborated the importance of structural elements by mutational analysis. This allowed identification of sequence motifs required for proper localization of PlaB and, interestingly, a NAD(H)-mediated tetramerization mechanism that controls activity and hence is suited to protect *L. pneumophila* from self-inflicted lysis.

6.1 Recombinant protein production of PlaB$_{D203N}$

All expressed constructs were generated and provided by Dr. Katja Kuhle and Dr. Wiebke Michel from the Robert Koch Institute, Wernigerode. PlaB$_{D203N}$, the PlaB version that carries a mutation in the active site was used for crystallization, was purified with an N-terminal Strep-tag and polished by SEC as shown in figure 6.1.

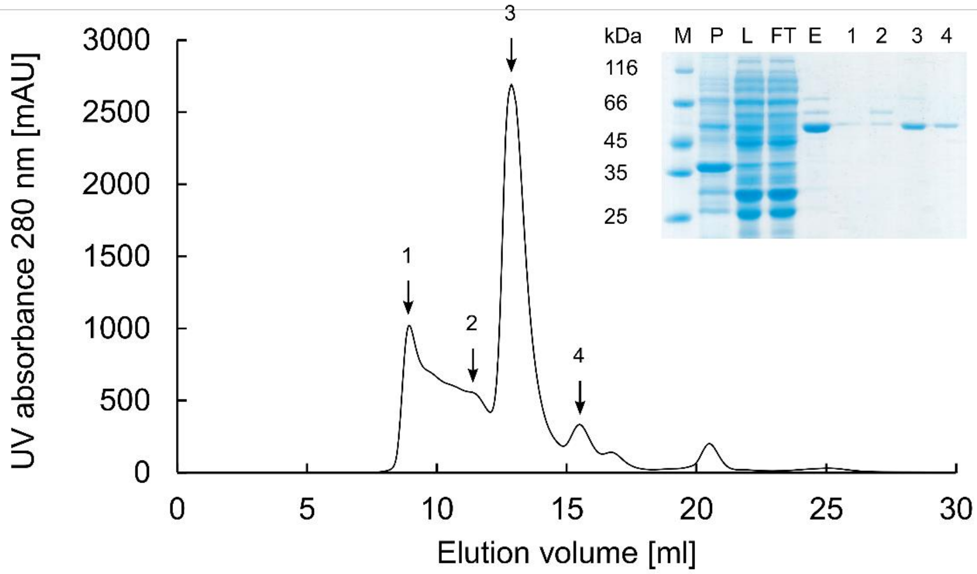


Figure 6.1: SEC on PlaB$_{D20N}$. Strep-tagged PlaB was expressed in *E. coli* BL21 cells. Cell lysate was applied on an 8 mL Strep-Tactin column and eluted. Pooled material was applied polished by SEC (S200 10/30 increase). Fractions were analyzed by SDS-PAGE. The main peak was pooled and further used for crystallization. M: Marker, P: cell pellet, L: cell lysate, FT: Strep-Tactin flowthrough, E: Strep-Tactin elution, 1-4: SEC fractions.

The purified protein was addressed by several biophysical techniques and subjected to crystallization.

6.2 Thermal shift assay with PlaB$_{D203N}$

In order to improve crystallization properties and diffraction quality, the buffer suitability for PlaB was addressed by a TSA. Buffer screening did not lead to any further improvement of the melting temperature. However, additive screening showed interesting alterations in the melting behavior upon addition of certain chemicals. The TSA assays were performed by mixing 5 µL protein solution at 2 mg/mL with 5 µL SYPRO® Orange (1:50 diluted from provided 5000x stock solution, Invitrogen, USA), 5 µL of 8-fold buffer (1200 mM NaCl and 400 mM Tris pH 8), and 35 µL RUBIC Additive Screen (Molecular Dimensions, USA) in a 96-well plate. Data was collected with a C1000 Touch™ Thermal Cycler, equipped with a CFX96™ Real-Time System from 4°C to 95°C in 1°C increments. TSA data was analyzed with the software BIO-RAD CFX96 Manager 3.1 (BIO-RAD, USA).

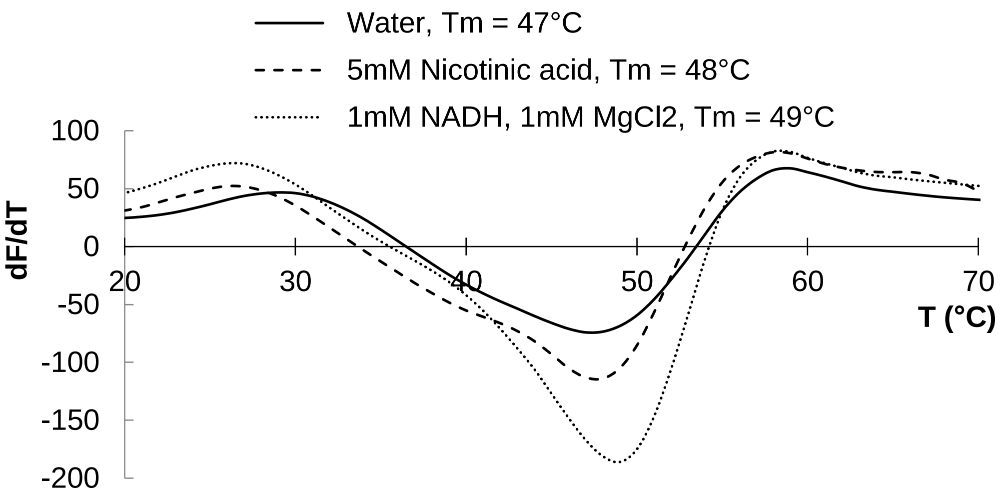


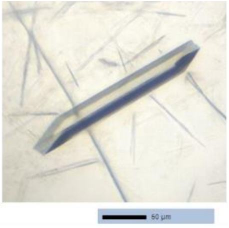
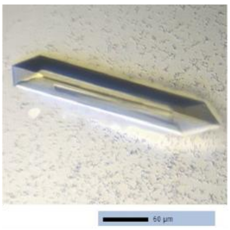
Figure 6.2: Thermal shift assay with PlaB$_{D203N}$. The results showed interesting behavior when nicotinic acid or NAD was added to the protein. Upon dilution with water, PlaB exhibits a broad melting signal starting at 35°C, ranging up to 58°C with a minimum at 47°C and a barely visible shoulder at approximately 40°C. When nicotinic acid was added to the protein, the two minima become more distinct. However, when NADH was added to PlaB, the minimum at 40°C is strongly reduced. At the same time, addition of NADH lead to a tremendous shift of melting signal towards the higher temperature melting minimum. The previously observed shoulder was almost completely diminished.

The curves in figure 6.2 show how the addition of nicotinic acid or NADH altered the melting behavior of PlaB. While the melting profile showed a flat curve with a minimum at 47°C and a shoulder at approximately 40°C for the reference, addition of NADH shifted the melting curve towards the species with the higher melting temperature. The shoulder on the other hand was drastically reduced. This indicated that addition of those chemicals lead to a more homogeneous sample (Chari et al., 2015). Based on general understanding, improvements not only of the melting temperature but also peak symmetry, leads to improved crystallization behavior (Boivin et al., 2013; Chari et al., 2015). Consequently, addition of the NAD^+-derivative SNAD was a strategy followed in the process of structural investigation.

6.3 Crystallization of $PlaB_{D203N}$

Extensive efforts were required to obtain high-quality crystals from PlaB. Most crystals suffered from strong anisotropic diffraction and streaky, overlapping reflections. Initial crystallization conditions were identified with the vapor diffusion method in sitting drops consisting of 200 nL protein solution at a concentration of 1 mg/mL to 6 mg/mL mixed with the same volume of precipitant and equilibrated against 60 µL reservoir at 20°C. Precipitants yielding crystals were optimized in grid and random screens. Microseeding experiments were performed with an OryxNano liquid dispenser (Douglas Instruments, UK), using 100 nL seed stock of native PlaB crystals broken up in 500 µL of the corresponding reservoir and 300 nL SeMet PlaB at 4 mg/mL mixed with 200 nL precipitant solution. At later stages, crystals were further optimized by pre-incubating the protein with 1 mM thio-nicotinamidedinucleotide (SNAD). Crystallization conditions used in this study are summarized in table 6.1.

Table 6.1: Crystallization conditions for PlaB$_{D203N}$. All batches were polished with PlaB SEC buffer. After extensive screening, crystals of highest quality were identified experimentally and selected for data collection. As crystallization was hampered by unknown reasons at first, seeding from a rod-shaped crystal, grown within 150 days from Oswald ripening, was applied. The seed stock was used to generate a single rod-shaped crystal from seleno-L-methionine derivatized PlaBD203N to solve the phase problem by experimental SAD-phasing. Another dataset was collected from crystals grown in an environment containing 1 mM thio-nicotinamidedinucleotide. Those crystals showed improved diffraction properties.

Crystal	Seed stock generation		Experimental phasing, $P2_1$		SNAD addition, P1	
Buffer	100 mM	HEPES pH 7	-	-	100 mM	Tris pH 8.5
Salt	18.9 mM	$LiSO_4$	77.8 mM	NaSNC	18.9 mM	LiSO4
Polymer	8.89% w/v	PEG mME 5k	5.44% v/v	PEG400	-	-
Alcohol	-	-	1.67% v/v	Glycerol	9.22% v/v	2-Propanol
Additive	5% v/v	Tacsimate pH7	4.11% v/v	Tacsimate pH7	1 mM	SNAD
	 Crystal developed by Oswald ripening within 150 days. Generation of seed stock		 Seleno-L-methionine derivatized crystal, used for experimental phasing. Generated from seed stock.		 Crystals grown after addition of 1 mM SNAD, enhanced diffraction qualities.	

Diffraction data collection proceeded with crystals obtained with precipitation conditions consisting of 100 mM Tris pH 8.5, 8.33% (v/v) glycerol, 10.8% (v/v) 2-propanol (native PlaB) and 77.8 mM NaSCN, 1.67% (v/v) glycerol, 5.44% (v/v) PEG400, 11.6% (v/v) 2-Propanol and 4.11% (v/v) Tacsimate (SeMet PlaB). After fishing, crystals were cryoprotected by washing in mother liquor supplemented with 10 % (v/v) (2*R*,3*R*)-(−)-2,3-butandiole before flash-cooling in liquid nitrogen. Diffraction data were collected on beamlines P11 (PETRAIII, DESY Hamburg, Germany) and PXIII (SLS, Paul Scherrer Institute, Villigen, Switzerland (Bingel-Erlenmeyer et al., 2011)). Indexing, integration and scaling was achieved with XDS and XSCALE (Kabsch, 2010 a; b), as summarized in table 6.2. The structure was finally determined at 2.5 Å resolution by single-wavelength anomalous dispersion experiment on a seleno-*L*-methionine labeled crystal obtained by seeding, collected at the K-absorption edge of selenium. This crystal belonged to space group $P2_1$ and the asymmetric unit contained four chains that arrange as a tetramer best described as a dimer of dimers. Heavy atom positions were identified with hkl2map (Pape et al., 2004) and forwarded to *phenix.phaser* (McCoy et al., 2007) of the PHENIX software suite (Adams et al., 2010). Initial model generation was performed with *phenix.autobuild* (Terwilliger et al., 2007) and supplemented by manual model building of poorly defined regions in Coot (Emsley et al., 2010) from the CCP4 software suite (Winn et al., 2011). Further refinement involved alternating rounds of manual adjustments and optimization in *phenix.refine* (Afonine et al., 2012). Model statistics are provided in table 6.2. High R-values are due to detected tNCS. The generated structure was further used as a MR model for another dataset, which was collected from a crystal that grew upon the addition of 1 mM SNAD. This crystal belonged to spacegroup P1 and showed good isotropy and well-defined peaks. No tNCS was detected for the data from cocrystallization with SNAD. Statistics are summarized in table 6.2

6.4 Crystal structure of PlaB$_{\mathbf{D203N}}$

PlaB crystallized in a tetrameric arrangement build from two identical dimers A and A’, as shown in figure 6.3. The PlaB monomer consists of two domains. An N-terminal domain (NTD) comprises a common α/β-hydrolase fold (ABH) (David et al., 1992; Schrag et al., 1997; Suplatov et al., 2012; Rauwerdink et al., 2015; Mindrebo et al., 2016), also found in other phospholipases (Rauwerdink et al., 2015). The internal β-sheet consists of 6 parallel β-strands (β2-β1-β3-β4-β5-β8 in the order of structural succession), antiparallelly complemented by β18 from a second PlaB molecule, forming a dimer bridging. The β-sheet is surrounded by a set of 17 α-helices (α1 to α17) representing the full ABH. The active site was determined previously as S85/D203/H251 (Bender et al., 2009) and was now structurally confirmed. It is embedded in the loops between the secondary structure elements β3-α5, α13-α14 and β8-α15 respectively. A lid structure (G127-G144; α8-α9), often found for proteins with an ABH (Tiesinga et al., 2007; Khan et al., 2017; Zhao et al., 2017), was found in a closed conformation. The NTD is followed by a unique C-terminal domain (CTD), starting at aa position Y325. It is terminated by the amino acid stretch R446 to D474 (in the following designated as hook). The CTD forms a bilobed β-sandwich that is complemented by intrusion of the non-canonical element β6/β7 from the NTD.

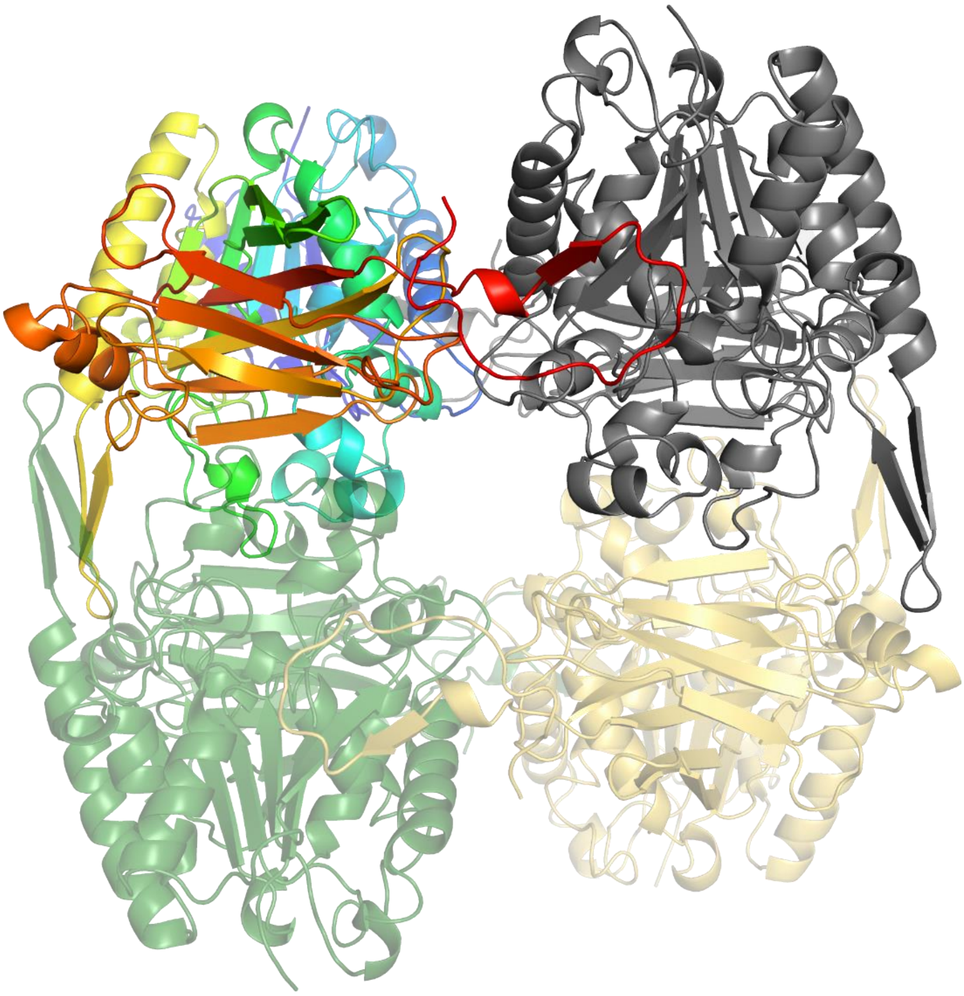


Figure 6.3: The crystal structure of PlaB$_{D203N}$. The tetrameric arrangement is depicted as a dimer of dimers. Monomers of dimer A are shown in a color gradient on the upper left and in grey on the upper right half. The dimer interface is mainly established by a C-terminal amino acid stretch, referred to as hook (red) that is attached to a second PlaB molecule. The second dimer is shown in the lower half of the image.

The hook extends the CTD and links two monomers to a dimer by interaction to the NTD of a second PlaB molecule and complements the ABH. The dimer interface was classified as stable by PISA-analysis with an energy gain of approximately -30 kcal/mol (Krissinel et al., 2007). Cation-π-interaction partners at the surface of the CTD decorate the tetramer interface and face each other in the center of the oligomer.

DALI search (Holm et al., 2016) resulted in a set of proteins which have a similar structure for the NTD from *Bacillus* ssp., *Geobacillus* spp., or *Streptomyces*, (1JI3 (Tyndall et al., 2002), 1KU0 (Jeong et al., 2002), 2DSN (Matsumara et al., 2008), 2Z5G (Matsumara et al., 2008), 3UMJ (Basri et al., 2012), 4X6U (Dror et al., 2015), 5H6B (Zhao et al., 2017)). None of those showed the secondary structure elements β6/β7, α17 and β9/β10 or the

CTD. Still it is a common feature of enzymes that the ABH is combined with a second domain for delivery of specific properties (Hopfner et al., 1998; Bashton et al., 2007).

Table 6.2: Data collection and refinement statistics PlaB$_{D203N}$. A SeMet derivatized crystal with spacegroup $P2_1$ was used for experimental phase determination in a SAD experiment. The generated model was further used as partial model for another dataset generated from a crystal after addition of 1 mM thio-nicotinamidedinucleotide in spacegroup P1.

Data collection	**SeMet structure**	**Native structure**
Space group	$P2_1$ (4)	P1 (1)
Cell dimensions		
a, *b*, *c* (Å)	75.8, 170.6, 93.5	75.8, 90.7, 92.8
a, b, g (°)	90.0, 92.9, 90.0	78.4, 90.2, 75.4
Wavelength (Å)	0.9795	1.0004
Resolution range (Å)	50.0 - 2.5 (2.6 – 2.5)	47.02 – 2.0 (2.1 – 2.0)
R_{meas}	22.5 (80.5)	18.5 (101.8)
I / s*I*	6.26 (2.12)	4.37 (1.04)
Completeness (%)	99.5 (99.7)	92.9 (83.0)
Redundancy	6.91 (6.99)	1.76 (1.48)
Refinement		
Resolution (Å)	2.5	2.0
No. reflections	81358	157093
R_{work} / R_{free}	0.27/0.32	0.22/0.27
No. atoms	16220	16341
Protein	15101	14916
Ligand/ion	229	424
Water	890	1001
B-factors		
Protein	33	22
Ligand/ion	30	22
R.m.s deviations		
Bond lengths (Å)	0.01	0.01
Bond angles (°)	1.25	1.23

The NTD has two non-canonical β-sheets (β6/β7, S216-R237 and β9/β10, K308-I321) and a non-canonical α-helix (α 17, R285-N305) which are exclusively found in PlaB from other

Legionella species but were absent in PlaB-like proteins from other organisms as identified by BLAST search, as shown in figure 6.4.

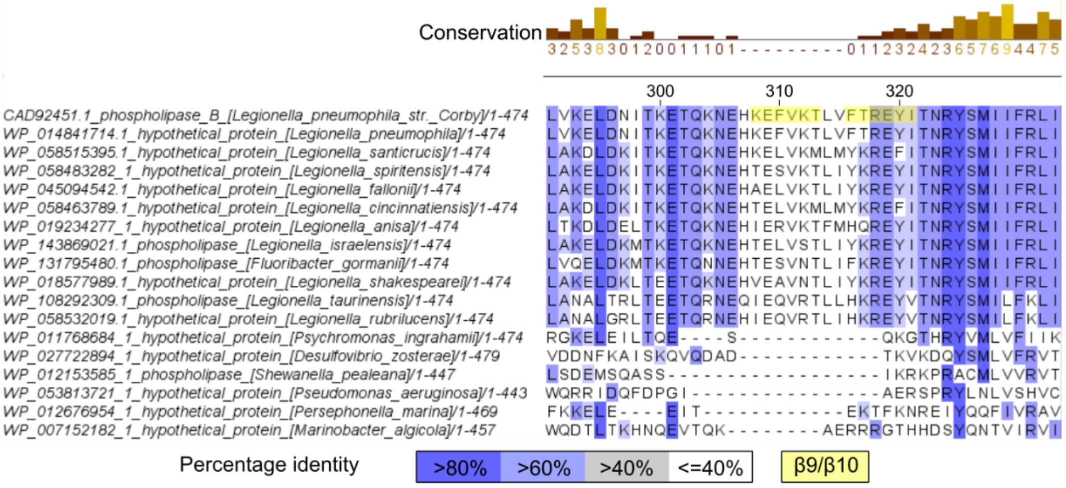


Figure 6.4: BLAST analysis of PlaB. Among close relatives of *L. pneumophila*, the non-canonical feature β9/β10(highlighted in yellow) is conserve and shows the characteristic cation-π-pattern. Homologs from other lung pathogens, such as *Pseudomonas aeruginosa* show a close overall homology to PlaB, however they lack the sheet β9/β10. BLAST similarity scores are listed in table S3.

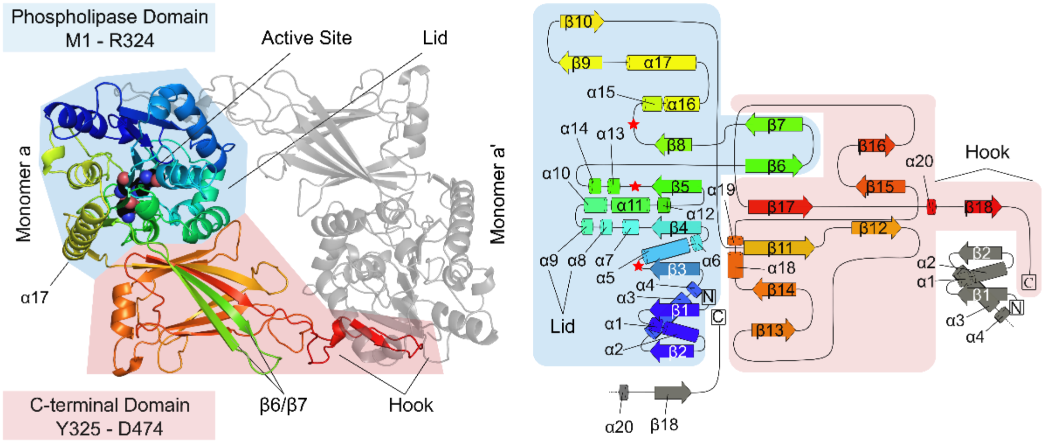


Figure 6.5: Structural architecture of PlaB. PlaB is divided in two domains, the N-terminal phospholipase domain (NTD, blue) and the C-terminal β-sandwich domain (CTD, red). The dimer (color gradient for monomer a and grey for monomer a', left) is shown. The active site is buried under a lid in closed conformation. The non-canonical sheet β6/β7 protrudes from the NTD and complements the CTD. The CTD evolves to a hook-like structure, that depicts the dimer interface. The topology diagram of a PlaB monomer (color gradient, right) exemplifies the α/β-hydrolase fold with the internal β-sheet (β2/β1/β3/β4/β5/β8 in the order of structural succession) and surrounding helices. The strand β18 is part of the hook-like dimer interface and complements the β-sheet in an antiparallel manner. The lid accommodates α8 and α9. The active site is depicted as red stars

A PlaB mutant single point mutation (S129A) from previous studies which turns out to be located in the lid structure showed reduced activity towards phosphatidyl choline (PC) but not phosphatidyl glycerol (PG) pointing towards its function in substrate recognition (Bender et al., 2009). Further surface exposed residues located in the lid structure were identified by structural investigations (S129, R130, R133). Mutations of these positions were addressed in subsequent experiments.

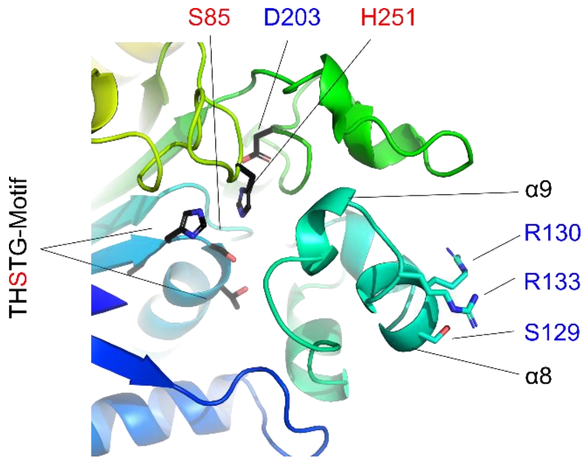


Figure 6.6: Lid structure of PlaB. The molecule has a lid structure, which is present in a closed conformation. It ranges from G127 to G144, covering α8 and α9. It effectively shields the active site S85, D203, H251 from the solvent. The non-canonical THSTG-Motif, used for classification of PlaB as the first member of a novel lipase family is shown, enclosing S85. Residues that were mutated for further analysis are labelled in blue.

The active site served for classification of PlaB as the first and so far, only member of a novel phospholipase family. No further phospholipases carry the THSTG-motif with the serine in the active site (Bender et al., 2009). The active site was effectively shielded from the solvent by a lid domain, found in a closed conformation. The lid domain was structurally determined to reach from amino acid position G127 to G144 and includes α8 and α9. In absence of substrates, lid domains are typically found in closed conformations (Khan et al., 2017).

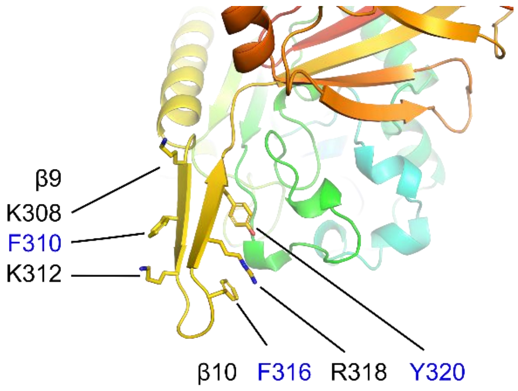


Figure 6.7: Non-canonical features of PlaB. The non-canonical structural feature β9/β10 (yellow) protrudes from the NTD. It is decorated with cation-π characters. Due to the antiparallel arrangement, a deletion mutant was generated. Also, the aromatic sidechains (highlighted in blue) were mutated to aspartates for further functional analysis.

Careful inspection of the non-canonical element β9/β10 reveals suspicious aa decoration with aromatic and cationic sidechains, as shown in figure 6.7. The exact role was addressed by PlaB versions lacking the π-character were subsequently characterized in terms of activity and localization.

The C-terminal domain (CTD) is shown in figure 6.8. This bilobed β-sandwich (upper lobe: β6-β7 intruded from NTD, β11-β13-β14-β17 and lower lobe: β12-β15-β16) shows an alternating, antiparallel/parallel/antiparallel composition in the upper lobe and antiparallel composition in the lower lobe. Inverse structural search for similar folds using DALI did not lead to any similar structure known, indicating that this C-terminal β-sandwich domain describes a novel fold (Holm et al., 2016; Holm, 2019).

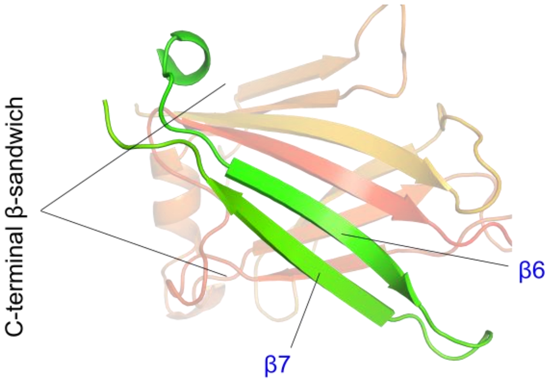


Figure 6.8: The sheet β6/β7 of PlaB (shown in green) protrudes from the NTD into the CTD (orange and red), which is a bilobed β-sandwich build from 7 β-strands, which are complemented by the β-sheet β6/β7.

Looking into closer detail, the CTD shows a high number of aromatic sidechains that are located in the tetramer interface. The exact function of those residues could not be clarified by generation of PlaB mutants. However, a role in membrane interaction seemed to be likely, as cationic arginine sidechains complement typical phospholipid interaction sites. Unpublished data shows that C-terminal truncation of PlaB did not affect membrane translocation until β12/β13 was deleted. Indeed, β12/β13 has a set of cation-π-interaction partners located on their surface, as shown in figure 6.9. This observation emphasizes a role of those residues in membrane translocation. However, further and more detailed investigations are needed to describe the exact role of the single residues.

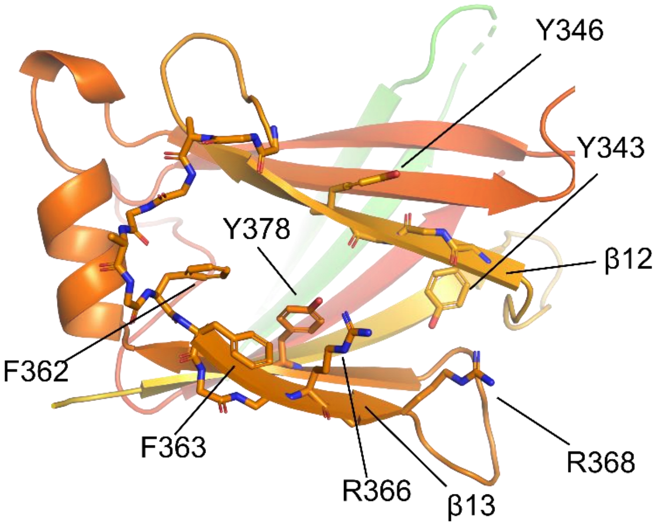


Figure 6.9: The C-terminal domain of PlaB is highly decorated with aromatic and cationic residues. Y343 and Y346 are located in β12, while F362, F363, R366 and R368 are found in β13. Y378 may play a role in a cation-π-interface and is found in β14.

In chapter 6.2 the influence of nicotinic acid and NAD$^+$ on the melting behavior of PlaB was described. Consequently, crystallization of PlaB with 1 mM of the NAD$^+$-derivative SNAD was performed as described in chapter 6.3. Crystallographic data quality was indeed improved by SNAD addition and tNCS was diminished. Inspection of the protein structure did not reveal any conformational changes upon SNAD addition. However, strong additional electron density could be observed in the tetramer interface, which was identified to originate from eight SNAD-molecules coordinated in the interface of the tetramer. This finding confirms the stabilization of a well-defined species, in this case the tetramer, by increasing TSA melting curve symmetry (Chari et al., 2015).

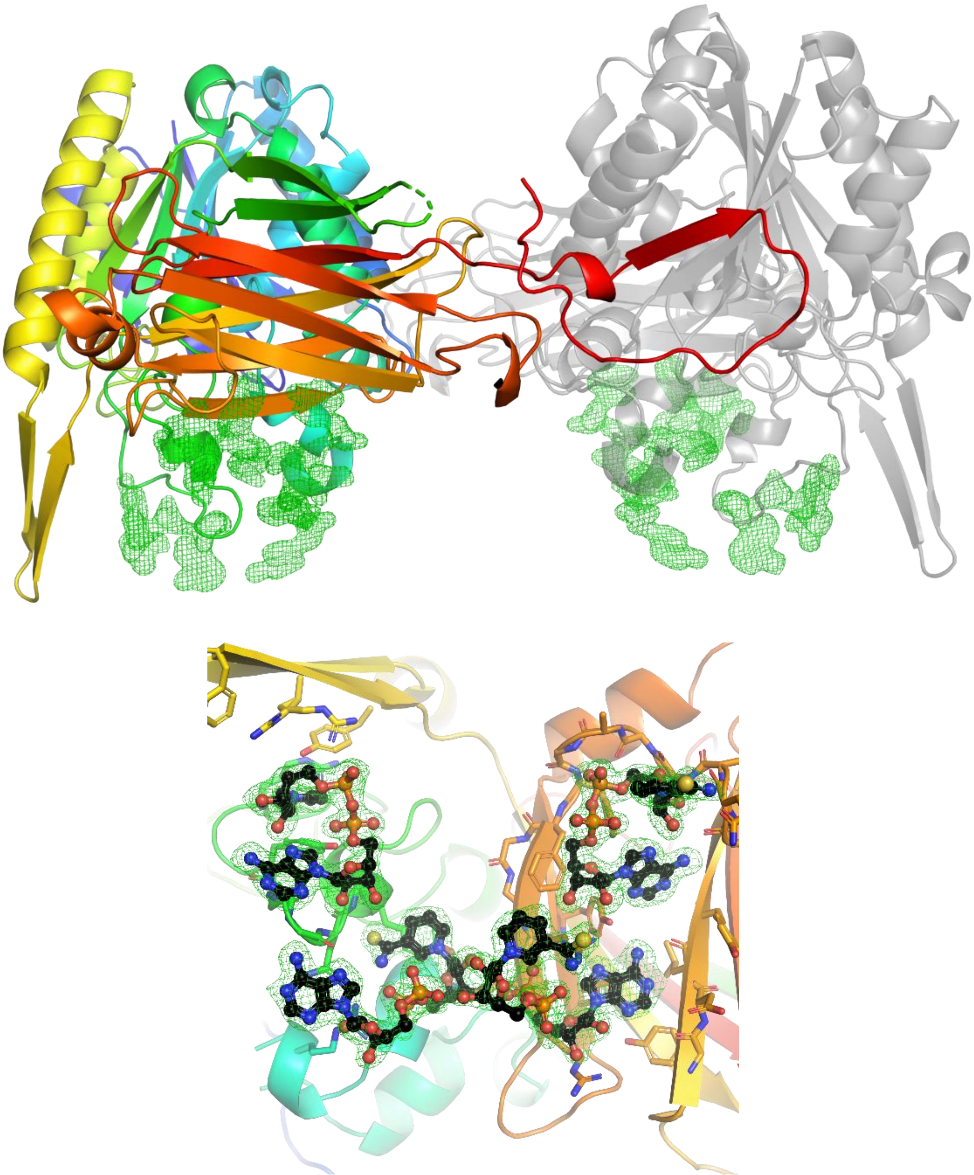


Figure 6.10: PlaB dimer in complex with SNAD. In the upper left is monomer A in a color gradient. In the upper right is monomer A´ in grey. Upon SNAD addition, uninterpreted fo-fc density emerged that was located in the tetramer interface (green mesh, σ = 3.0). The density was interpreted to originate from added SNAD. Four molecules are shown in the interface of monomer A and A’ (black carbons, blue nitrogens, red oxygens, orange phosphorus, yellow sulfurs). The interactions involve β9/β10 (yellow), a loop region and helix α13 (green), the lid (cyan) and β12/β13 located in the CTD (orange).

The localization of SNAD in the tetramer interface by binding and stabilization of diverse structural elements suggested that a close cross-play of oligomerization, inhibition and localization of PlaB exists. Given that activity was known to be inhibited by tetramerization, the reverse conclusion that tetramer-promoting nucleotides inhibit PlaB activity was addressed by measuring activity of PlaB version in absence and presence of nucleotides.

6.5 PlaB activity assays

Mutant generation and activity assays were performed by Dr. Wiebke Michel, Dr. Christina Lang and Christian Galisch. Structure guided mutagenesis was performed to shed light on the mode of oligomerization and activity. Bacterial lysates of *E. coli* BL21 producing PlaB or PlaB mutants were incubated with dipalmitoylphosphatidylcholine (PC), dipalmitoylphosphatidylglycerol (PG), 1-monopalmitoyl-lysophosphatidylcholine (LPC), and 1-monopalmitoyl-lysophosphatidylglycerol (LPG) (Avanti Polar Lipids, Inc.). The release of fatty acids was determined using NEFA-HR(2) Assay (Wako Fujifilm) as described previously (Bender et al., 2009). Figure 6.11 shows the activity of the PlaB versions towards the different substrates.

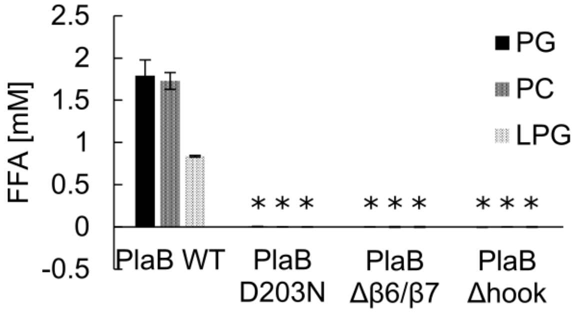


Figure 6.11: PlaB mutant activity assays. Activity assays performed with cell lysates with wildtype PlaB or different PlaB mutants (WT, inactive mutant D203N, Δβ6/β7 and Δhook) after adding the substrates PG, PC and LPC. PlaB WT showed activity towards all substrates. The mutants show no activity. Error bars indicate standard deviations (SD). Statistical analysis was performed using two-tailed, unpaired students t-test, relating PlaB WT to PlaB mutants. * $p < 0.02$ (N=3).

PlaB versions D203N, Δβ6/β7 and Δhook did not show any residual enzymatic activity. D203N served as a negative control with a mutation silencing PlaB activity, as described before (Bender et al., 2009; Kuhle et al., 2014). Deletion of β6/β7 showed the same reduction and emphasizes the importance of this unique element. Presumably, β6/β7 serves as an anchor, holding the CTD in a certain orientation relative towards the NTD and therewith enables PlaB to form homodimers, mediated by the hook structure. Consequently, hook deletion led to comparable results. This confirms the previously observed importance of the CTD and parts thereof for catalytic activity (Bender et al., 2009; Kuhle et al., 2014).

Previously, a single point mutant S129A was under investigation (Bender et al., 2009). With the presented crystal structure, we can confirm the changed substrate specificity is based on the residue’s role in substrate recognition. While activity against phosphatidylglycerols was slightly but significantly reduced, phosphatidylcholines could not be turned over any more. This points towards substrate recognition of the positively charged choline by interaction with the hydroxyl group of S129. Structural analysis unveiled an exposed position of S129 in the lid structure. The sidechains of R130 and

R133 were also identified to present charged and surface exposed residues and were selected for structure guided mutagenesis. Results are shown in figure 6.12.

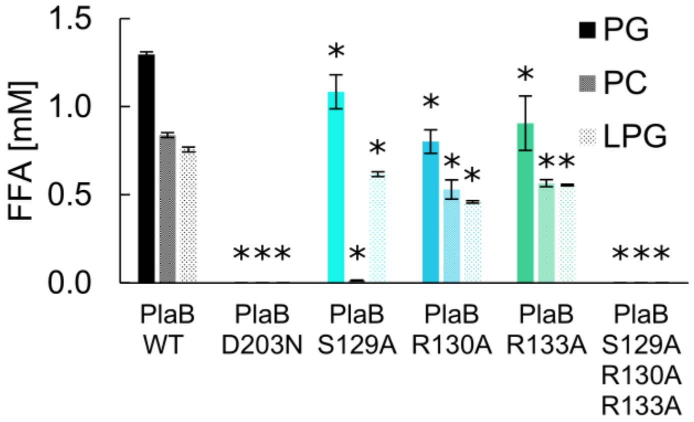


Figure 6.12: PlaB activity assays, lid domain. Activity assays performed with cell lysates for PlaB versions (WT, inactive mutant D203N and lid mutations S129A, R130A and R133A, as well as a triple mutant). The color code refers to the lid as structural element to be mutated (compare figure 6.6). The S129A version lost activity towards cholesterol carrying PC. R130A and R133A show slightly reduced overall activity. The triple mutant lost activity towards all substrates tested. Error bars indicate SD. Statistical analysis was performed using two-tailed, unpaired students t-test, relating PlaB WT to PlaB versions. * $p < 0.02$ (N=3).

The S129A mutant could not efficiently hydrolyze PC. Further, single point mutations in the lid structure (R130A, R133A) showed reduced activity for all substrates tested, pointing towards a more common role in substrate recognition. Mutation of the cluster S129A/R130A/R133A showed no remaining activity. These experiments underline the relevance of the amino acids in the lid for substrate recognition, as already described in different studies (Holmquist et al., 1995; Secundo et al., 2004; Schleberger et al., 2007; Khan et al., 2017). Interestingly, R130 exhibits an ionic interaction to a coordinated SNAD molecule in the tetramer interface. Presumably, this mechanism ensures that the lid domain remains in a closed conformation as long as the PlaB tetramer is present.

Another structural element of particular interest is β9/β10, as this non-canonical element showed a high degree of surface-exposed cationic and aromatic side chains, typically found in membrane interacting protein regions (Matsumara et al., 2008; Bale et al., 2010; Grauffel et al., 2013). Therefore, a Δβ9/β10 deletion mutant and a F310D/F316D/Y320D triple mutant were generated and tested on their activity, as shown in figure 6.13. Both versions showed reduced activities for all substrates tested.

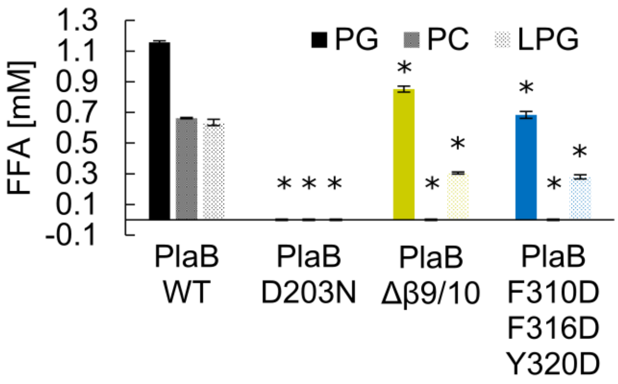


Figure 6.13: PlaB activity assays, β9/β10. Activity assays performed on purified protein samples for PlaB versions (WT, inactive mutant D203N, Δβ9/β10 and F310D/F316D/Y320D). Error bars indicate SD. Statistical analysis was performed using two-tailed, unpaired students t-test, relating PlaB WT to PlaB versions. * $p < 0.02$ (N=3).

Both PlaB versions showed a similar pattern in activity reduction, indicating that the aromatic side chains participate in the turnover of the substrates. The residues were mutated to aspartates instead of alanines, as the negatively charged side chains of the acid groups were meant to interact with neighboring arginines and to diminish the aromatic character at the same time. In β9/β10, different types of interactions are involved in the binding of SNAD molecules that promote tetramerization. Y320 shows a typical interaction of its hydroxyl group with the heteroaromatic pyridine ring system of the SNAD molecule, which is also coordinated by Y192 from the other side via π-stacking. Especially this binding site is likely causing an effect in melting behavior for nicotinic acid (compare chapter 6.2), which is mimicking the nicotinamide group of NAD(H).

As β9/β10 contributes to SNAD coordination and the tetramer interface, the enzymatic activities of PlaB versions were compared in presence of excess NAD^+ or NADH, respectively. Results are shown in figure 6.14.

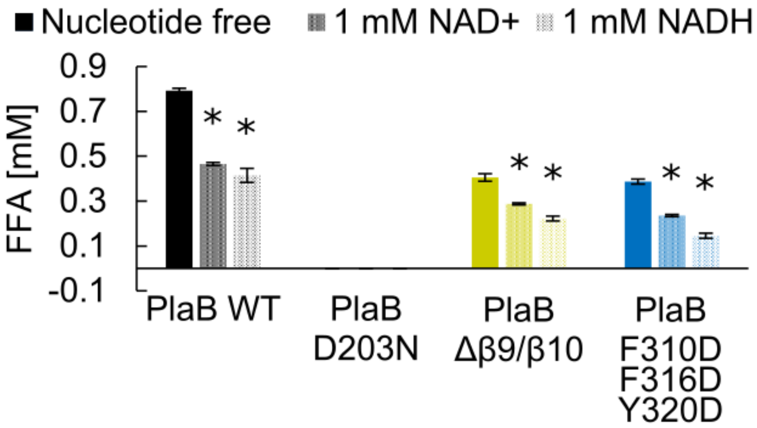


Figure 6.14: PlaB activity assays, nucleotide substitution. Activity assays performed with purified protein samples of PlaB versions (WT, D203N, Δβ9/β10 and F310D/F316D/Y320D) with and without 1 mM NAD^+ and NAD(H) respectively. Error bars indicate SD. Statistical analysis was performed using two-tailed, unpaired students t-test, relating PlaB versions without supplemented nucleotides to PlaB versions with 1 mM NAD^+ and NAD(H), respectively. * $p < 0.02$ (N=3).

This experiment showed a strong reduction in PlaB WT activity upon addition of NAD$^+$ or NAD(H), presumably due to enhanced oligomerization mediated inhibition. The activities of PlaB versions Δβ9/β10 and F310D/F316D/Y320D were in general reduced compared to the WT. The oligomerization state in dependency of nucleotide presence was addressed by SEC-MALS experiments.

6.6 SEC-MALS with PlaB

PlaB was described to show concentration dependent oligomerization, accompanied by enzymatic inhibition. At low concentrations in the range below 10 nM, the protein was present in an active form with a MW corresponding to a dimer. At 1 µM, PlaB was strongly inhibited and present as a tetramer (Kuhle et al., 2014). During protein purification dedicated to crystallization experiments, only high concentrations of PlaB were handled. Therefore, almost exclusively tetrameric protein was observed during polishing SEC, as one can see in figure 6.15.

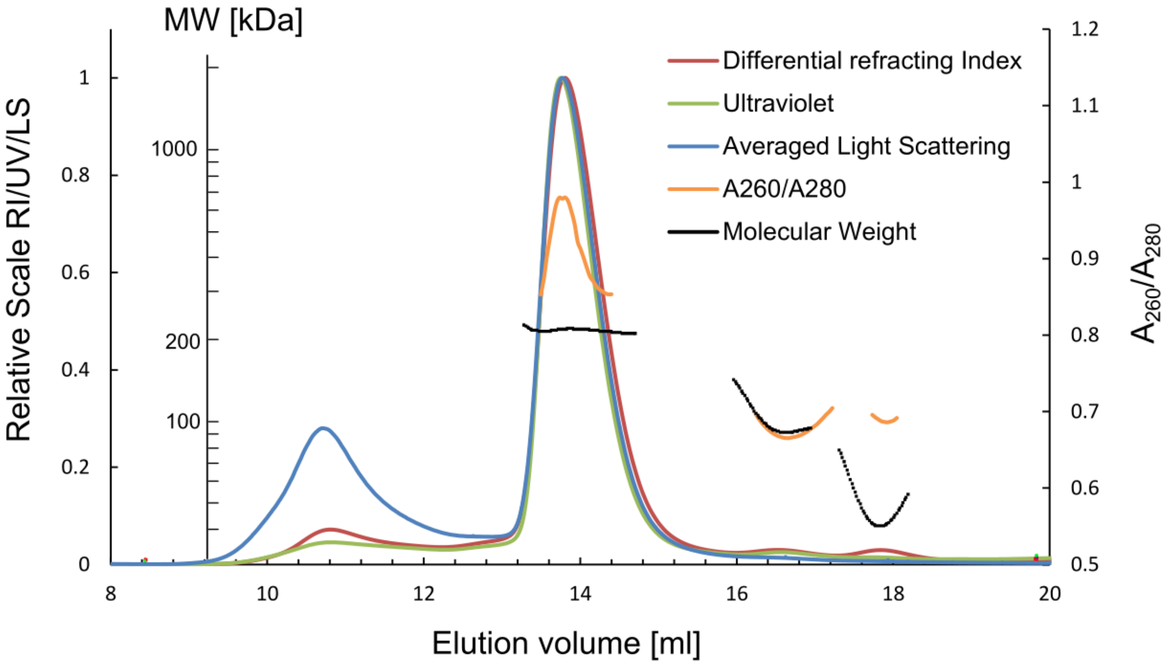


Figure 6.15: SEC-MALS results for PlaB$_{D203N}$. Here, 10 mg of PlaB$_{D203N}$ were applied in parallel runs on a GE ÄKTA system to monitor the absorption at 280 nm and 260 nm and the light scattering on a Wyatt SEC-MALS system. The elution profiles were aligned and superimposed. The tetramer was determined with a mass of 228 kDa (13.4 – 14.8 mL). Interestingly, the A$_{260}$/A$_{280}$ ratio increases to more than 0.9, indicating nucleotide binding. This was not true for dimeric (118 kDa, 16.1 – 17.4 ml) and monomeric (62 kDa, 17.4 – 18.3 mL) species that were present in minor fractions. Here, the ratio numbered to 0.7.

For the tetramer an increased ratio for the absorptions at 260 nm and 280 nm was observed, pointing towards nucleotide binding of the tetramer, as it could be suggested from structural inspections (see figure 6.10) and enzymatic inhibition assays (see figure 6.14). This observation was only partly true for dimeric and monomeric species, that were present to much lesser extent. However, as the crystal structure unveiled β9/β10 as a non-canonical feature, present only in close PlaB homologes, structure guided mutagenesis was applied to deepen understanding of the exact function. From the model, involvement in tetramerization, NAD^+-coordination and membrane association could be derived already. As the sheet is of antiparallel nature, a deletion mutant could be generated. The deletion mutant, as well as the triple mutant F310D/F316D/Y320D showed a similarly and significantly reduced activity towards all substrates tested (see figure 6.13). In another experiment, the inhibitory properties of the nucleotides NAD^+ and NADH were observed for the PlaB WT but also for PlaB versions affecting β9/β10. Consequently, the oligomeric state of PlaB β9/β10 versions investigated by SEC-MALS, as summarized in figure 6.16.

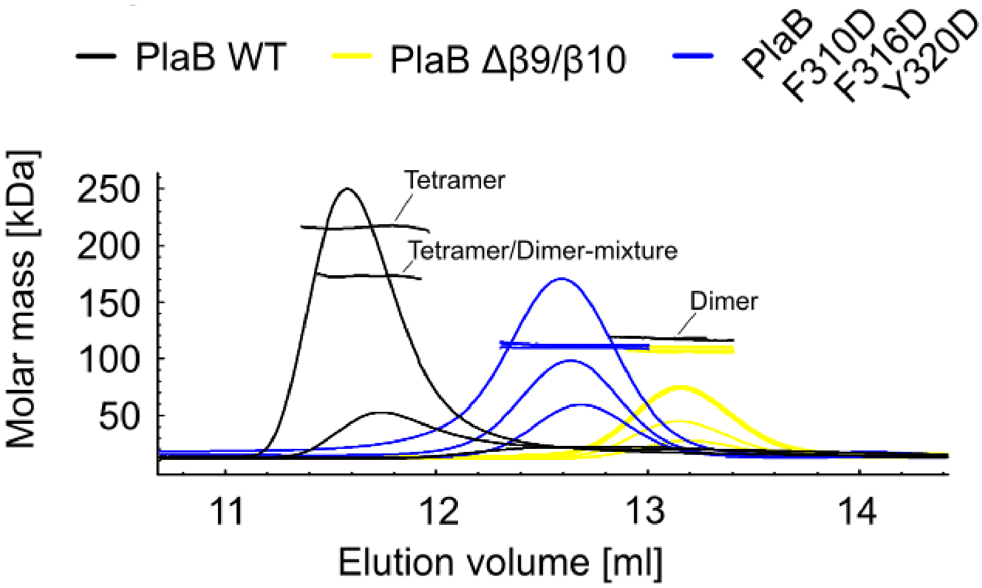


Figure 6.16: SEC-MALS results for PlaB mutants. SEC-MALS showed concentration dependent oligomerization of PlaB WT (black). PlaB is present almost exclusively as a tetramer (120 µg), as a mixture of tetramers and dimers (60 µg) or exclusively as dimer (30 µg). PlaB mutants Δβ9/β10 and F310D/F316D/Y320D tested on their oligomerization were applied with the same amounts. The mutants showed exclusive dimer formation, independent from the applied amount.

Various amounts of protein were applied to monitor the oligomerization state in dependency of the local concentration of PlaB. As one can see from figure 6.16, the PlaB WT (black) showed a concentration-dependent behavior. The tetramer with a MW of 216.4 kDa was found when 120 µg of PlaB were injected. Here, the local PlaB concentration at the peak was determined to be 1.94 µM. Injection of 60 µg showed strongly reduced light scattering signal for a tetramer/dimer mixture with a calculated mixed mass of 173.3 kDa. Here, a second dimer peak was observed, resulting in a MW of 117.6 kDa. The local PlaB peak concentrations were determined to be 0.41 µM and 0.12 µM, respectively. The injection of 30 µg of PlaB only resulted in PlaB dimers with a calculated MW of 118.7 kDa and a peak concentration of 0.17 µM. These results are in line with observations made before, that showed PlaB in an inactive tetrameric state at a concentration above 1 µM (Kuhle et al., 2014). The β9/β10 and F310D/F310D/Y320D PlaB mutants on the other hand did not exhibit concentration-dependent oligomerization. Instead, regardless of the the amount of protein sample applied, exclusively dimers were observed. Local PlaB peak concentrations were determined with 1.02 µM, 0.57 µM and 0.26 µM for Δβ9/β10 (yellow) and 2.77 µM, 1.34 µM and 0.81 µM for F310D/F310D/Y320D (blue). Results are summarized in table 6.3. Activity assays have shown inhibition upon addition of 1 mM NAD(H) (compare figure 6.14). As activity was linked to oligomerization, the question was addressed, whether the presence of 1 mM NAD$^+$ affects tetramerization. Therefore, the SEC-MALS system was equilibrated with SEC-MALS buffer, supplemented with NAD$^+$. The resulting chromatograms are summarized in figure 6.17.

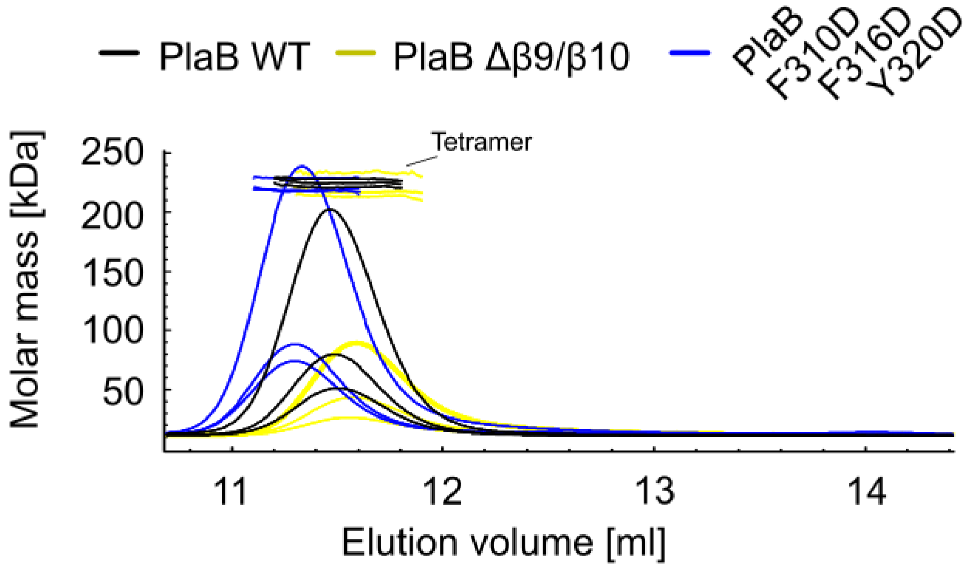


Figure 6.17: SEC-MALS results for PlaB mutants, supplemented with nucleotides. SEC-MALS experiments on PlaB WT (black, 60 µg, 30 µg and 15 µg), and structure guided mutants (PlaB Δβ9/β10, yellow and F319D/F312D/Y320D, blue) were exclusively present as tetramers after addition of 1 mM NAD$^+$ to the SEC buffer. All species were almost exclusively present as tetramers, regardless of applied amounts of protein (60 µg, 30 µg and 15 µg) and PlaB version.

Indeed, addition of 1 mM NAD$^+$ promoted PlaB tetramerization, as it had been predicted from structural and biochemical data gathered before. Interestingly, the concentration dependency for PlaB WT (black) was lost and exclusively tetramers with a MW of 225.1 kDa ± 1.2% were observed. The local PlaB peak concentrations were determined to be 1.12 µM, 0.44 µM and 0.24 µM. The β9/β10 mutants showed slightly differing elution volumes, indicating changes in the gyrical radius of the tetramer, but calculated masses were almost identical to the WT with 221.6 kDa ± 3.9% and 222.3 kDa ± 2.2%, respectively. The local PlaB peak concentrations for the mutants were determined with 0.48 µM, 0.20 µM and 0.09 µM for Δβ9/β10 and 1.52 µM, 0.51 µM and 0.37 µM for F310D/F316D/Y320D. Results are summarized in table 6.3.

Direct comparison of SEC-MALS experiments of PlaB WT and β9/β10 versions with and without supplemented NAD$^+$, complement the figure drawn before. It can be concluded that concentration dependent tetramerization occurred when no nucleotide was supplemented. Concentrations of PlaB of approximately 2 µM were able to bind copurified nucleotides with a higher efficiency, compared to lower concentrations. The results confirm previous studies (Kuhle et al., 2014). The β9/β10 versions lost their ability to hold copurified nucleotides and were present as dimers. Still, they had the ability to form tetramers, when NAD$^+$ was present in large excess.

Table 6.3: Summary SEC-MALS results for PlaB variants. Results for the WT, Δβ9/β10 and F310D/F316D/Y320D and effect of NAD$^+$ on tetramerization are shown. In absence of NAD$^+$, PlaB WT forms concentration dependent tetramers and dimers. Δβ9/β10 and F310D/F316D/Y320D were exclusively found as dimers. Upon addition of 1 mM NAD$^+$ to the SEC buffer, exclusively tetramers were found for all proteins and concentrations tested.

Sample	applied amount [µg]	regular SEC buffer		applied amount [µg]	SEC buffer + 1 mM NAD$^+$
		Dimer MW (kDa) / c (µM)	**Tetramer** MW (kDa) / c (µM)		**Tetramer** MW (kDa) / c (µM)
PlaB WT	120	-	216.4 / 1.94	60	228.3 / 1.12
	60	117.6 / 0.12	173.3 / 0.41	30	225.3 / 0.44
	30	118.7 / 0.17	-	15	221.8 / 0.24
	mean MW	-	-	mean MW	225.1 ± 1.2%
PlaB Δβ9/β10	120	107.8 / 1.02	-	60	217.5 / 0.48
	60	111.0 / 0.44	-	30	213.6 / 0.20
	30	106.3 / 0.24	-	15	233.7 / 0.09
	mean MW	108.4 ±1.8%	-	mean MW	221.6 ± 3.9%
PlaB F310D/F316D/Y320D	120	110.2 / 2.77	-	60	218.7 / 1.52
	60	113.1 / 1.34	-	30	219.1 / 0.51
	30	112.2 / 0.81	-	15	229.2 / 0.37
	mean MW	111.8 ± 1.1%	-	mean MW	222.3 ± 2.2%

Structural interpretation of the PlaB oligomer interfaces by PISA showed interesting properties. The hook was found to stabilize the dimer with 29.8 kcal/mol and the dimer was classified as a stable complex. For the tetramer, the conditions are more complex. When SNAD was included in the calculations, the tetramer was calculated to be stabilized by 45 kcal/mol. However, when SNAD was removed from the model structure, almost no remaining energy gain was predicted for the tetramerization (Krissinel et al., 2007). This suggests, that the PlaB dimer is the catalytically active form of PlaB. The non-canonical β9/β10 sheet seems to be a multifunctional element that promotes nucleotide dependent oligomerization and plays an additional role in enzymatic activity, regardless of the oligomeric state. The inhibition in the tetrameric state arises from nucleotide binding, as SNAD was found to lock the lid in its closed conformation by ionic interaction.

6.7 Cell fractionation and PlaB localization

Despite several years of research, the translocation mechanism of PlaB to the outer membrane remains unclear. Deletion mutants suggest that the transport must be ensured either by an unknown export machinery or by an auto transport process. It has been found that transport of C-terminal truncated PlaB was unaffected until β12/β13 was deleted. However, deletion of β12/β13 lead to abrupt loss of export ability (data not published). Inspection of the crystal structure showed that explicitly those strands show a set of cation-π-interaction sites and tightly interact with coordinated SNAD molecules. Beside those located in β12/β13, further aromatic and cationic sidechains were observed to be surface exposed in the CTD. Those residues have been selected for structure guided mutagenesis. To prevent the denaturation of PlaB versions upon modification, target amino acids were selected carefully and the direct environment was taken into account.

Cloning and cell fractionation were performed by Dr. Wiebke Michel at the RKI, Wernigerode. The separation of the inner and outer membranes of *E. coli* BL21 carrying different vectors was performed with changes according to Roy & Isberg (Filip et al., 1973; Roy et al., 1997). Results for subcellular fractionation are summarized in figure 6.18.

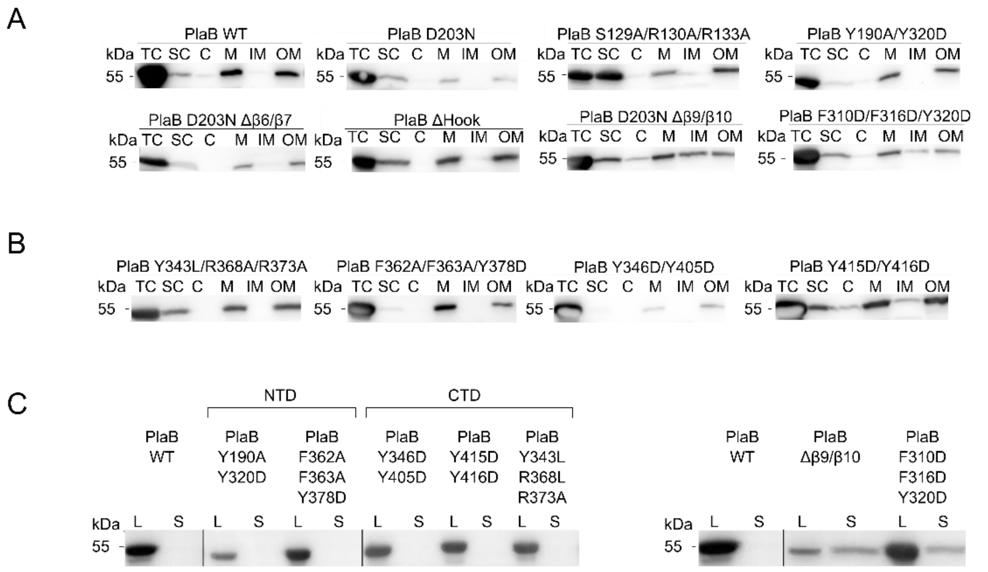


Figure 6.18: Western blot analysis on PlaB variants. WB was performed using an anti-Strep-tag antibody. It shows localization of strep-tagged PlaB versions after subcellular fractionation. A) PlaB versions with mutations in the active site, the lid domain, the noncanonical elements β6/β7 and β9/β10 as well as the hook domain were analysed with respect to their localization. All versions were soluble. In contrast to any other version, deletion of β9/β10 and modification in this element lead to the presence of PlaB at the inner membrane. B) PlaB version that carried modifications in the CTD did not show any striking difference in membrane localization. C) Culture supernatant was tested on PlaB presence. When modifications affected β9/β10, PlaB was additionally found in the culture supernatant. TC: total content, SC: soluble content, C: cytosol, M: membrane, IM: inner membrane, OM: outer membrane, L: culture lysate, S: culture supernatant.

Fractions were analyzed by means of Western blotting, detecting PlaB (StrepMAB-Classic, HRP conjugate, Iba), the outer membrane control protein OmpA (antibody kindly provided by Prof. Prasadaro Nemani (Prasadarao et al., 1996), the inner membrane control protein LepB (antibody kindly provided by Prof. Heinje (De Buck et al., 2007, 2008; Vranckx et al., 2007) and the cytosolic control protein DnaK (Enzo Life Sciences).

Comparing the presence of PlaB in the different subcellular fractions as shown in figure 6.18, one can see presence of soluble protein, presumably possible by PlaB tetramer formation. Changes in the amount of detected soluble PlaB could potentially arise from higher expression levels, enabling a higher fraction to form tetramers. All versions tested were also present in the membrane fraction. However, more detailed analysis and separate examinations showed that PlaB was originally present only at the outer membrane. Changes were observed when β9/β10 was affected by mutagenesis. For those versions, PlaB was also found at the inner membrane. The same versions did also show reduced ability to bind to the OM, as PlaB could be detected in the cell culture supernatant. Once more independent experiments emphasize the importance of β9/β10 for PlaB specific characteristics, such as concentration dependent tetramerization and even membrane translocation and interaction. Contrary to expectations, mutation of the aromatic residues in β12/β13 (Y343L/R368/R373A and F362A/F363A/Y378D) did not show any effect on PlaB localization. However, PlaB version in which the two neighbored tyrosins Y415D and Y416D were mutated exhibit a similar localization pattern to PlaB Δβ9/β10. The tyrosins interact with R44 and R447 and are located directly in the transition area of CTD and hook domain. The exact role of those residues in transportation, activity and oligomerization still needs to be elaborated.

6.8 Summary - Phospholipase PlaB

To unravel the molecular basis for enzyme oligomerization, inhibition and localization, the crystal structure of PlaB was generated in its tetrameric form. The PlaB tetramer is a dimer of identical dimers. The monomer consists of an N-terminal phospholipase domain with an ABH-fold augmented by two non-canonical β-sheets, β6/β7 and β9/β10. The CTD reveals a novel fold displaying a bilobed β-sandwich with a hook structure that is required for dimerization and complementation of the N-terminal phospholipase domain in the neighboring molecule. Unexpectedly, we observed additional electron density in the dimer/dimer-interface, suggesting that the bound nucleotides stabilize the tetramer and hence lead to inactivation of the enzyme. The density was interpreted to arise from NAD(H). Indeed, addition of NAD^+, NADH or SNAD, an NAD^+ analogs to PlaB, led to increased melting, enhanced crystallization, reduced activity and promoted tetramerization. The deletion of β9/β10, which acts as a nucleotide interaction site, altered oligomerization, activity profiles and localization, pointing towards a multifunctional role of this non-canonical element. Observations made were summarized in figure 6.19.

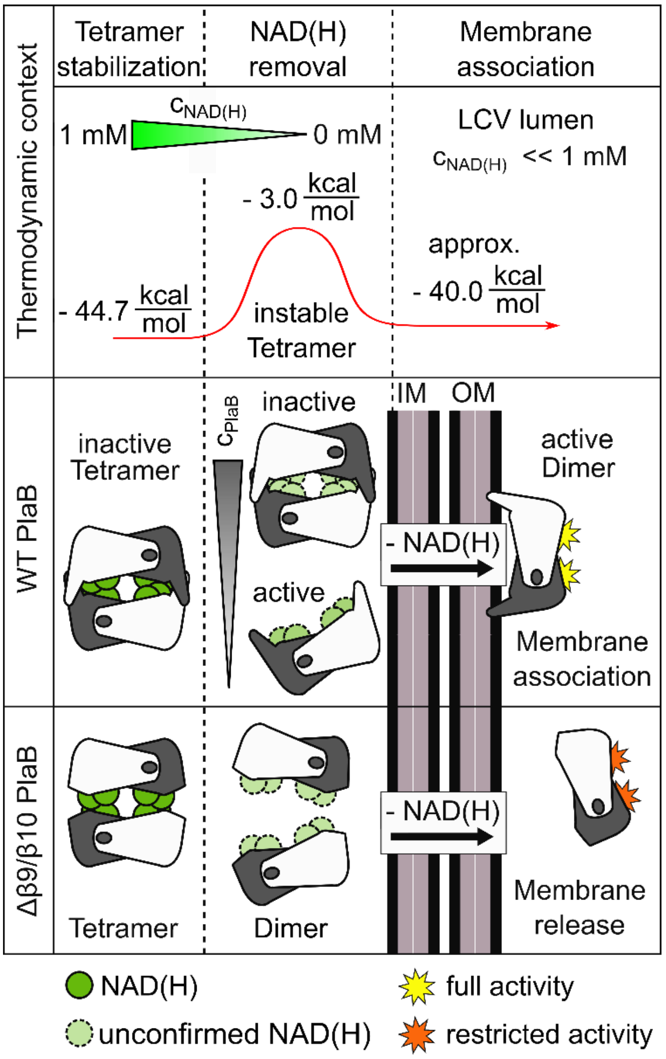


Figure 6.19: Summary of results for PlaB. NAD(H) plays a critical role in β9/β10-mediated and concentration dependent oligomerization and activity regulation. In the presence of NAD(H), the tetramer is stabilized by approximately 45 kcal/mol, which results in exclusive tetramer formation when nucleotides are present in excess. This is even true for the Δβ9/β10 mutant. High concentrations of intracellular NAD(H) mediate tetramer formation of PlaB and thus inhibits the lipolytic activity. Upon translocation into a nucleotide-free environment, PlaB tetramers dissociate concentration dependently into dimers, as the tetramer interface is now only stabilized by 3 kcal/mol, which is considered as an unstable interface. The importance of β9/β10 for the tetramerization was shown, as mutants were exclusively present as dimers, when nucleotides were absent. Upon decomposition of the tetramer, β9/β10 becomes accessible for membrane association, which will gain approximately 40 kcal/mol for two dimers with two β9/β10-elements each upon insertion. Therefore, dimers will associate to the outer membrane of *L. pneumophila* and will be presented as the active form. NAD(H) is shown as green spheres. Unclear binding of NAD(H) at PlaB dimers is shown as light green sphere with dashed contour. An accessible catalytic centre is illustrated as yellow star. The restricted activity for the Δβ9/β10 version is illustrated as orange star.

Together, these data reveal a novel ligand-mediated oligomerization-dependent inhibition mechanism that likely protects *L. pneumophila* from self-inflicted damage by PlaB until the enzyme is exported to the bacterial surface. PlaB forms tetramers, build from two dimers and is activated by tetramer deoligomerization. Unexpectedly, we found that inactive tetramers are stabilized by NAD(H) molecules bound to the dimer/dimer interface. Since the cytosolic concentration of NAD(H) is in the millimolar range whereas it drops to zero outside of the cell, we propose that this unprecedented NAD(H)-dependent oligomerization mechanism protects *L. pneumophila* from damage until PlaB is exported.

The dimer on the other hand was found to be stabilized by 29.8 kcal/mol, independently from nucleotide concentrations, which was considered as stable. The antiparallel complementation of the sheet at β2 by β-strand β18 from a second PlaB molecule is a feature that has not been observed before and indicates the importance of dimerization for enzyme activity.

7 Conclusion

This thesis covered several distinct projects. Firstly, the proline reductase complex, a central metabolic key player in the nosocomial pathogen *C. difficile*, was addressed and the assembly of a hetero-triacontamer and its subunits PrdAα/β and PrdB were studied. Secondly, the crystal structures of the two phospholipases PlaA and PlaB, important virulence factors from the airborne *L. pneumophila,* were determined that extend our current understanding of their mode of action. Although the pathogens included in this thesis were highly divergent, experienced complications in capturing concrete protein oligomerization states were alike.

For the proline reductase supercomplex, the N-terminal section of PrdA provides a complicated decamerization platform of unknown architecture. This is introducing high flexibility and a dynamic oligomerization equilibrium paired with a large molecular weight, making protein crystallization for structural elucidation inaccessible. The interaction of PrdA and PrdB is preserved, when PrdA is N-terminally truncated. The natively purified proline reductase complex demanded chemical crosslinking and could successfully be visualized at low resolution. Combination of results from various techniques showed a complicated assembly from 30 subunits into a supercomplex of 970 kDa and 30 nm in diameter, exhibiting a 5-fold symmetry. For detailed understanding however, a more comprehensive structural characterization, providing higher resolution, will be necessary. After all, the feasibility of different techniques in structural biology generated encouraging results for different sets of subunits. Cryo-EM on $\text{PrdA}_{149\text{-}626}\text{PrdB}_{\text{U151C}}$ resulted in an anisotropic map with local resolution of approximately 4 Å. Negative stain TEM on the triacontameric proline reductase supercomplex has shown good particle distribution and repetitive patterns in picked particles, however, class averaging suffered from low resolution due to high intrinsic flexibility. Last, NMR spectroscopy on $\text{PrdB}_{\text{U151C}}$ was shown to provide chemical shifts of high resolution but $\text{PrdB}_{\text{U151C}}$ was present in a mixture of at least two species with different conformations. Therefore, reproducible sample preparation needs to be improved. This could possibly be achieved by discovery of unknown cofactors involved in oligomerization or alteration of the chemical environment. The presented results provide a robust ground for ongoing studies. In the future, a major focus should be set on identification of possible cofactor binding and purification reproducibility. Conceivable methods for cofactor analysis are MS after protein denaturation and cofactor release for organic molecules or ICP-MS for ions.

The second project in this thesis was the structural elucidation of the phospholipases PlaA and PlaB from *L. pneumophila*. PlaA depicts a compact and mostly rigid 34 kDa monomer resulting in protein crystals of high quality. Structural elucidation based on experimental phase determination by iodine soaking was successful and confirmed sequence-based models, as well as previous knowledge about a common ABH-fold and the active site location. New insights were gained with respect to the change of substrate specificity upon ProA-mediated PlaA maturation. It has been found that a 29 amino acid long sequence, framed by a cysteine bridge, serves as a lid domain, which shields the active site from the solvent. At the same time, a putative binding site for the cholesterol moiety of the alternative substrate cholesterol palmitate was identified to be accessible on the surface of PlaA. Proteolytic maturation of the lid domain might be responsible for induced flexibility in the ABH-fold. This allows binding of the actual substrate LPC (16:0) and covering the cholesterol palmitate binding site, finally triggering PLA activity. The structure will help to generate structure guided PlaA versions, which will potentially provide deeper understanding of binding modes and substrate specificities. Until now, co-crystallization of PlaA with the enzymatic product palmitic acid was highly beneficial. However, co-crystallization with substrates such as LPC (16:0) and cholesterol palmitate was not successful.

PlaB was characterized by a complicated oligomerization behavior, possibly arising from a domain without sequence-based structural models. Concentration dependent oligomerization had been reported before. Diverging results in batch purifications provoked poor reproducibility in crystallization and mostly, collected datasets suffered from strong crystallographic defects, such as tNCS. However, it was found that, contrary to existent understanding, PlaB tetramerization was not solely concentration dependent but rather nucleotide mediated. Supplementation of NAD$^+$ mediated concentration independent tetramer formation and hence improved crystallization of the PlaB tetramer. The crystal structure of a PlaB dimer of dimers explains the importance of the C-terminal domain, exhibiting a novel fold and in particular the 15 C-terminal amino acids are necessary for dimerization and activity. This short sequence, referred to as hook, serves as a tight and stable dimer interface and complements the ABH-fold of the second PlaB molecule. Tetramerization is based on a more complicated cross play of nucleotides and surface-exposed cation-π-interaction sites. One particular, multifunctional element was identified, the antiparallel β-strand β9/β10, which is involved in enzymatic activity, nucleotide binding, tetramerization and membrane association. Additional cation-π-characters in the tetramer interface might contribute to the unique use of up to eight molecules of nucleotides for oligomerization, as future experiments will clarify. Bound nucleotides were responsible for enzymatic inhibition that was accompanying tetramerization, as ionic interactions between the PlaB lid domain and a nucleotide presumably stabilized the lid in its closed conformation. This hypothesis requires more detailed investigation, still. But the influence of surface exposed charged sidechains in the lid domain on substrate recognition and turnover could unambiguously be monitored already. The enigmatic mechanism of PlaB export lacks complete understanding, still. However, unpublished data paired with structural inspections suggest that beside β9/β10, also β12/β13 may play a critical role in transmembrane transport. Altogether, structural determination of PlaB uncovered the architecture of the C-terminal domain with a novel fold and clearly improved understanding of dimerization dependent activity and tetramerization induced inhibition. It is suggested that for the purpose of self-protection, intracellular PlaB may be present as an inactive tetramer due to a high level of nucleotides. Upon membrane translocation, the tetrameric state is transformed to an active and thereby cytotoxic dimeric state, accompanied by nucleotide removal. This results in the exposure of PlaB on the bacterial surface of *L. pneumophila*. Until now, no comparable mode of regulation has been reported (Afonine et al., 2012).

8 References

Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., 2010: PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*., **66**, 213–221.

Adamus, K.; Le, S. N.; Elmlund, H.; Boudes, M.; Elmlund, D., 2019: AgarFix: Simple and accessible stabilization of challenging single-particle cryo-EM specimens through crosslinking in a matrix of agar. *Journal of Structural Biology*., **207**, 327–331.

Afonine, P.; Klaholz, B.; Moriarty, N.; Poon, B.; Sobolev, O.; Terwilliger, T.; Adams, P.; Urzhumtsev, A., 2018: New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallographica Section D: Structural Biology*., **74**, 814–840.

Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., 2012: Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography*., **68**, 352–367.

Akoh, C. C.; Lee, G. C.; Liaw, Y. C.; Huang, T. H.; Shaw, J. F., 2004: GDSL family of serine esterases/lipases. *Progress in Lipid Research*., **43**, 534–552.

Albert, A.; Dhanaraj, V.; Genschel, U.; Khan, G.; Ramjee, M. K.; Pulido, R.; Sibanda, B. L.; Delft, F. Von; Witty, M.; Blundell, T. L.; Smith, A. G.; Abell, C., 1998: Crystal structure of aspartate decarboxylase at 2.2 A resolution provides evidence for an ester in protein self-processing. *Nature structural biology*., **5**, 289–293.

Andreesen, J. R., 2004: Glycine reductase mechanism. *Current Opinion in Chemical Biology*., **8**, 454–461.

Awad, M. M.; Johanesen, P. A.; Carter, G. P.; Rose, E.; Lyras, D., 2014: Clostridium difficile virulence factors: insights into an anaerobic spore-forming pathogen, **5**, 579–593.

Babnigg, G.; Jedrzejczak, R.; Nocek, B.; Stein, A.; Eschenfeldt, W.; Stols, L.; Marshall, N.; Weger, A.; Wu, R.; Donnelly, M.; Joachimiak, A., 2015: Gene selection and cloning approaches for co-expression and production of recombinant protein–protein complexes. *Journal of Structural and Functional Genomics*., **16**, 113–128.

Bale, S.; Ealick, S. E., 2010: Structural Biology of S-Adenosylmethionine Decarboxylase, **38**, 451–460.

Banerji, S.; Aurass, P.; Flieger, A., 2008: The manifold phospholipases A of Legionella pneumophila – Identification , export , regulation , and their link to bacterial virulence. *International Journal of Medical Microbiology*., **298**, 169–181.

Barkin, J. A.; Sussman, D. A.; Fifadara, N.; Barkin, J. S., 2017: Clostridium difficile Infection and Patient-Specific Antimicrobial Resistance Testing Reveals a High Metronidazole Resistance Rate. *Digestive Diseases and Sciences*., **62**, 1035–1042.

Bartlett, J. G., 1979: Antibiotic-Associated Pseudomembranous Colitis. *Reviews of Infectious Diseases*., **1**, 530–539.

Bartlett, J. G., 2002: Clostridium difficile-associated Enteric Disease. *Current Infectious Disease Reports 2002,*., **4**, 477–483.

Bashton, M.; Chothia, C., 2007: The Generation of New Protein Functions by the Combination of Domains. *Structure*., **15**, 85–99.

Basri, M.; Ruslan, R.; Leow, T. C.; Salleh, A. B.; Ali, M. S. M.; Rahman, R. N. Z. R. A., 2012: Improvement of Thermal Stability via Outer-Loop Ion Pair Interaction of Mutated T1 Lipase from Geobacillus zalihae Strain T1. *International Journal of Molecular Sciences*., **13**, 943–960.

Battaglioli, E. J.; Hale, V. L.; Chen, J.; Jeraldo, P.; Ruiz-mojica, C.; Schmidt, B. A.; Rekdal, V. M.; Till, L. M.; Huq, L.; Smits, S. A.; Moor, J.; Jones-hall, Y.; Smyrk, T.; Khanna, S.; Pardi, D. S.; Patel, R.; Chia, N.; Nelson, H.; Sonnenburg, J. L., 2018: Clostridioides difficile uses amino acids associated with gut microbial dysbiosis in a subset of patients with diarrhea. *Science Translational Medicine*., **10**, 1–27.

Bednarski, B.; Andreesen, J. R.; Pich, A., 2001: In vitro processing of the proproteins GrdE of protein B of glycine reductase and PrdA of D-proline reductase from Clostridium sticklandii: Formation of a pyruvoyl group from a cysteine residue. *European Journal of Biochemistry*., **268**, 3538–3544.

Bender, J.; Rydzewski, K.; Broich, M.; Schunder, E.; Heuner, K.; Flieger, A., 2009: Phospholipase PlaB of Legionella pneumophila represents a novel lipase family. Protein residues essential for lipolytic activity, substrate specificity and hemolysis. *Journal of Biological Chemistry*., **284**, 27185–27194.

Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., 2000: The Protein Data Bank www.rcsb.org. *Nucleic acids research*., **28**, 235–242.

Bingel-Erlenmeyer, R.; Olieric, V.; Grimshaw, J. P. A.; Gabadinho, J.; Wang, X.; Ebner, S. G.; Isenegger, A.; Schneider, R.; Schneider, J.; Glettig, W.; Pradervand, C.; Panepucci, E. H.; Tomizaki, T.; Wang, M.; Schulze-Briese, C., 2011: SLS crystallization platform at beamline X06DA-A fully automated pipeline enabling in situ X-ray diffraction screening. *Crystal Growth and Design*., **11**, 916–923.

Blanchet, C. E.; Spilotros, A.; Schwemmer, F.; Graewert, M. A.; Kikhney, A.; Jeffries, C. M.; Franke, D.; Mark, D.; Zengerle, R.; Cipriani, F.; Fiedler, S.; Roessle, M.; Svergun, D. I., 2015: Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *Journal of Applied Crystallography*., **48**, 431–443.

Bloch, F., 1946: Nuclear induction. *Physical Review*., **70**, 460–474.

Boivin, S.; Kozak, S.; Meijers, R., 2013: Optimization of protein purification and characterization using Thermofluor screens. *Protein Expression and Purification*., **91**, 192–206.

Böttcher, B.; Wynne, S. A.; Crowther, R. A., 1997: Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*., **386**, 88–91.

Böttcher, B.; Nassal, M., 2018: Structure of Mutant Hepatitis B Core Protein Capsids with Premature Secretion Phenotype. *Journal of Molecular Biology*., **430**, 4941–4954.

Bouillaut, L.; Self, W. T.; Sonenshein, A. L., 2013: Proline-dependent regulation of Clostridium difficile stickland metabolism. *Journal of Bacteriology*., **195**, 844–854.

Bouillaut, L.; Dubois, T.; Sonenshein, A. L.; Dupuy, B., 2015: Integration of metabolism and virulence in Clostridium difficile. *Research in Microbiology*., **166**, 375–383.

Bouillaut, L.; Dubois, T.; Francis, M. B.; Daou, N.; Monot, M.; Sorg, J. A.; Sonenshein, A. L.; Dupuy, B., 2019: Role of the global regulator Rex in control of NAD + -regeneration in Clostridioides (Clostridium) difficile, **111**, 1671–1688.

Bragg, W. H., 1913: The reflection of X-rays by Crystals. *Proceedings of the royal society A*., **88**, 428–438.

Buchan, D. W. A.; Jones, D. T., 2019: The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic acids research*., **47**, W402–W407.

Buck, M., 2003: Crystallography: Embracing conformational flexibility in proteins. *Structure*., **11**, 735–736.

Carter, G. P.; Rood, J. I.; Lyras, D., 2012: The role of toxin A and toxin B in the virulence of Clostridium difficile. *Trends in Microbiology*., **20**, 21–29.

Case, D. A., 1998: The use of chemical shifts and their anisotropies in biomolecular structure determination. *Current Opinion in Structural Biology*., **8**, 624–630.

Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Skelton, N. J., 2007: *Protein NMR Spectroscopy: Principles and Practice*. Elsevier Academic Press.

Chari, A.; Haselbach, D.; Kirves, J. M.; Ohmer, J.; Paknia, E.; Fischer, N.; Ganichkin, O.; Möller, V.; J Frye, J.; Petzold, G.; Jarvis, M.; Tietzel, M.; Grimm, C.; Peters, J. M.; Schulman, B. A.; Tittmann, K.; Markl, J.; Fischer, U.; Stark, H., 2015: ProteoPlex: Stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. *Nature Methods*., **12**, 859–865.

Chen, J.; Noble, A. J.; Kang, J. Y.; Darst, S. A., 2019: Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *Journal of Structural Biology: X*., **1**, 100005.

Chu, F.; Thornton, D.; Nguyen, H., 2019: Chemical cross-linking in the structural analysis of protein assemblies Feixia. *Methods*., **144**, 53–63.

Clamp, M.; Cuff, J.; Searle, S. M.; Barton, G. J., 2004: The Jalview Java alignment editor. *Bioinformatics*., **20**, 426–427.

Cohen, S. H.; Gerding, D. N.; Johnson, S.; Kelly, C. P.; Loo, V. G.; McDonald, L. C.; Pepin, J.; Wilcox, M. H., 2010: Clinical practice guidelines for Clostridium difficile infection in adults: 2010 update by the society for healthcare epidemiology of america (SHEA) and the infectious diseases society of America (IDSA). *Infection control and hospital epidemiology*., **31**, 431–455.

Collins, M. D.; Lawson, P. A.; Willems, A.; Cordoba, J. J.; Fernandez-Garayzabal, J.; Garcia, P.;

Cai, J.; Hippe, H.; Farrow, J. A. E., 1994: The phylogeny of the genus Clostridium: Proposal of five new genera and eleven new species combinations. *International Journal of Systematic Bacteriology*., **44**, 812–826.

Conover, G.; Martinez-Morales, F.; Heidtman, M.; Luo, Z. Q.; Tang, M.; Chen, C.; Geiger, O.; Isberg, R., 2008: Phosphatidylcholine synthesis is required for optimal function of Legionella pneumophila virulence determinants. *Cell Microbiology*., **10**, 514–528.

Cornely, O. A.; Nathwani, D.; Ivanescu, C.; Odufowora-Sita, O.; Retsa, P.; Odeyemi, I. A. O., 2014: Clinical efficacy of fidaxomicin compared with vancomycin and metronidazole in Clostridium difficile infections: A meta-analysis and indirect treatment comparison. *Journal of Antimicrobial Chemotherapy*., **69**, 2892–2900.

Creasey, E. A.; Isberg, R. R., 2012: The protein SdhA maintains the integrity of the Legionella-containing vacuole. *Proceedings of the National Academy of Sciences of the United States of America*., **109**, 3481–3486.

Davanloo, P.; Rosenberg, A. H.; Dunn, J. J.; Studier, F. W., 1984: Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*., **81**, 2035–2039.

David, L.; Cheah, E.; Cygler, M.; Dijkstra, B.; Frolow, F.; Sybille, M.; Harel, M.; James Remington, S.; Silman, I.; Schrag, J.; Joel, L.; Koen, H. G. V.; Goldman, A., 1992: The α/β hydrolase fold. *Protein Engineering, Design and Selection*., **5**, 197–211.

De Buck, E.; Vranckx, L.; Meyen, E.; Maes, L.; Vandermissen, L.; Anne, J.; Lammertyn, E., 2007: The twin-arginine translocation pathway is necessary for correct membrane insertion of the Rieske Fe/S protein in Legionella pneumophila. *FEBS Letters*., **581**, 259–264.

De Buck, E.; Höper, D.; Lammertyn, E.; Hecker, M.; Anné, J., 2008: Differential 2-D protein gel electrophoresis analysis of Legionella pneumophila wild type and Tat secretion mutants. *International Journal of Medical Microbiology*., **298**, 449–461.

Dembek, M.; Barquist, L.; Boinett, C. J.; Cain, A. K.; Mayho, M.; Lawley, T. D.; Fairweather, N. F.; Fagan, R. P., 2015: High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *mBio*., **6**, e02383-14.

Denisov, I. G.; Sligar, S. G., 2016: Nanodiscs for structural and functional studies of membrane proteins. *Nature Structural and Molecular Biology*., **23**, 481–486.

Dror, A.; Kanteev, M.; Kagan, I.; Gihaz, S.; Shahar, A.; Fishman, A., 2015: Structural insights into methanol-stable variants of lipase T6 from Geobacillus stearothermophilus. *Applied Microbiology and Biotechnology*., **99**, 9449–9461.

Drulyte, I.; Johnson, R. M.; Hesketh, E. L.; Hurdiss, D. L.; Scarff, C. A.; Porav, S. A.; Ranson, N. A.; Muench, S. P.; Thompson, R. F., 2018: Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallographica Section D: Structural Biology*., **74**, 560–571.

Dubochet, J.; Adrian, M.; Lepault, J.; McDowall, A., 1985: Emerging techniques: Cryo-electron microscopy of vitrified biological specimens. *Trends in Biochemical Sciences*., **10**, 143–146.

Dubochet, J.; Adrian, M.; Chang, J.; Lepault, J.; McDowall, A., 1987: Cryoelectron Microscopy of Vitrified Specimens. *Cryotechniques in Biological Electron Microscopy*., 114–131.

Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N., 2005: Flexible nets: The roles of intrinsic disorder in protein interaction networks. *FEBS Journal*., **272**, 5129–5148.

Dupuy, B.; Sonenshein, A. L., 1998: Regulated transcription of Clostridium difficile toxin genes. *Molecular Microbiology*., **27**, 107–120.

Durfee, T.; Nelson, R.; Baldwin, S.; Plunkett, G.; Burland, V.; Mau, B.; Petrosino, J. F.; Qin, X.; Muzny, D. M.; Ayele, M.; Gibbs, R. A.; Csörgo, B.; Pósfai, G.; Weinstock, G. M.; Blattner, F. R., 2008: The complete genome sequence of Escherichia coli DH10B: Insights into the biology of a laboratory workhorse. *Journal of Bacteriology*., **190**, 2597–2606.

Ekstrom, J. L.; Tolbert, W. D.; Xiong, H.; Pegg, A. E.; Ealick, S. E., 2001: Structure of a human S-adenosylmethionine decarboxylase self-processing ester intermediate and mechanism of putrescine stimulation of processing as revealed by the H243A mutant. *Biochemistry*., **40**, 9495–9504.

Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., 2010: Features and development of Coot . *Acta Crystallographica Section D Biological Crystallography*., **66**, 486–501.

Ferraris, R. P.; Yasharpour, S.; Lloyd, K. C. K.; Mirzayan, R.; Diamond, J. M., 1990: Luminal glucose concentrations in the gut under normal conditions. *American Journal of Physiology - Gastrointestinal and Liver Physiology*., **259**.

Filip, C.; Fletcher, G.; Wulff, J. L.; Earhart, C. F., 1973: Solubilization of the cytoplasmic membrane of Escherichia coli by the ionic detergent sodium lauryl sarcosinate. *Journal of*

Bacteriology., **115**, 717–722.

Finnerty, W. R.; Makula, R. A.; Feeley, J., 1979: Cellular Lipids of the Legionnaires' Disease Bacterium. *Annals of Internal Medicine*., **90**, 631–634.

Flieger, A.; Neumeister, B.; Cianciotto, N. P., 2002: Characterization of the Gene Encoding the Major Secreted Lysophospholipase A of Legionella pneumophila and Its Role in Detoxification of Lysophosphatidylcholine. *Infection and Immunity*., **70**, 6094–6106.

Flieger, A.; Rydzewski, K.; Banerji, S.; Broich, M.; Heuner, K., 2004: Cloning and characterization of the gene encoding the major cell-associated phospholipase a of Legionella pneumophila, plaB, exhibiting hemolytic activity. *Infection and Immunity*., **72**, 2648–2658.

Fliermans, C. B.; Cherry, W. B.; Orrison, L. H.; Smith, S. J.; Tison, D. L.; Pope, D. H., 1981: Ecological distribution of Legionella pneumophila. *Applied and Environmental Microbiology*., **41**, 9–16.

Flores-Díaz, M.; Monturiol-Gross, L.; Naylor, C.; Alape-Girón, A.; Flieger, A., 2016: Bacterial Sphingomyelinases and Phospholipases as Virulence Factors. *Microbiology and Molecular Biology Reviews*., **80**, 597–628.

Florin, I.; Thelestam, M., 1991: ADP-ribosylation in Clostridium difficile toxin-treated cells is not related to cytopathogenicity of toxin B. *Biochimica et biophysica acta*., **1**, 51–54.

Fonknechten, N.; Chaussonnerie, S.; Tricot, S.; Lajus, A.; Andreesen, J. R.; Perchat, N.; Pelletier, E.; Gouyvenoux, M.; Barbe, V.; Salanoubat, M.; Paslier, D. Le; Weissenbach, J.; Cohen, G. N.; Kreimeyer, A., 2010a: Clostridium sticklandii , a specialist in amino acid degradation : revisiting its metabolism through its genome sequence.

Fonknechten, N.; Chaussonnerie, S.; Tricot, S.; Lajus, A.; Andreesen, J. R.; Perchat, N.; Pelletier, E.; Gouyvenoux, M.; Barbe, V.; Salanoubat, M.; Le Paslier, D.; Weissenbach, J.; Cohen, G. N.; Kreimeyer, A., 2010b: Clostridium sticklandii, a specialist in amino acid degradation:revisiting its metabolism through its genome sequence. *BMC genomics*., **11**, 555.

Franco, I. S.; Shuman, H. A.; Charpentier, X., 2009: The perplexing functions and surprising origins of Legionella pneumophila type IV secretion effectors. *Cellular Microbiology*., **11**, 1435–1443.

Franke, D.; Petoukhov, M. V; Konarev, P. V; Panjkovich, A.; Tuukkanen, A. T.; Mertens, H. D. T.; Kikhney, A.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; Svergun, D. I., 2017: ATSAS 2.8 : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography*., **50**, 1212–1225.

Gallaghert, T.; Rozwaski, D. A.; Ernst, S. R.; Hackert, M. L., 1992: Refined Structure of the Pyruvoyl-dependent Histidine Decarboxylase from Lactobacillus 30a. *Journal of Molecular Biology*., **230**, 516–528.

Garapati, S.; Monteith, W.; Wilson, C.; Kostenko, A.; Kenney, J. M.; Danell, A. S.; Burns, C. S., 2018: Zn2+-binding in the glutamate-rich region of the intrinsically disordered protein prothymosin-alpha. *Journal of Biological Inorganic Chemistry*., **23**, 1255–1263.

Ghanim, G. E.; Kellogg, E. H.; Nogales, E.; Rio, D. C., 2019: Structure of a P element transposase–DNA complex reveals unusual DNA structures and GTP-DNA contacts. *Nature Structural and Molecular Biology*., **26**, 1013–1022.

Gomez-Valero, L.; Rusniok, C.; Carson, D.; Mondino, S.; Pérez-Cobas, A. E.; Rolando, M.; Pasricha, S.; Reuter, S.; Demirtas, J.; Crumbach, J.; Descorps-Declere, S.; Hartland, E. L.; Jarraud, S.; Dougan, G.; Schroeder, G. N.; Frankel, G.; Buchrieser, C., 2019: More than 18,000 effectors in the Legionella genus genome provide multiple, independent combinations for replication in human cells. *Proceedings of the National Academy of Sciences of the United States of America*., **116**, 2265–2273.

Grant, S. G. N.; Jessee, J.; Bloom, F. R.; Hanahan, D., 1990: Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proceedings of the National Academy of Sciences of the United States of America*., **87**, 4645–4649.

Grauffel, C.; Yang, B.; He, T.; Roberts, M. F.; Gershenson, A.; Reuter, N., 2013: Cation-π interactions as lipid-specific anchors for phosphatidylinositol-specific phospholipase C. *Journal of the American Chemical Society*., **135**, 5740–5750.

Greenfield, N. J., 2009: Using circular dichroism collected as a funcion of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature protocols*., **1**, 2527–2535.

Gurtler, V., 1993: Typing of Clostridium difficile strains by PCR-amplification of variable length 16S-23S rDNA spacer regions. *Journal of General Microbiology*., **139**, 3089–3097.

Hales, L. M.; Shuman, H. A., 1999: Legionella pneumophila contains a type II general secretion pathway required for growth in amoebae as well as for secretion of the Msp protease. *Infection* *and Immunity*., **67**, 3662–3666.

Hall, I.; O'toole, E., 1935: Intestinal flora in new-born infants with a description of a new pathogenic anaerobe, Bacillus difficili. *American Journal of Diseases of Children*., **49**, 390.

Hammitt, M. C.; Bueschel, D. M.; Keel, M. K.; Glock, R. D.; Cuneo, P.; DeYoung, D. W.; Reggiardo, C.; Trinh, H. T.; Songer, J. G., 2008: A possible role for Clostridium difficile in the etiology of calf enteritis. *Veterinary Microbiology*., **127**, 343–352.

Hauer, F.; Gerle, C.; Fischer, N.; Oshima, A.; Shinzawa-Itoh, K.; Shimada, S.; Yokoyama, K.; Fujiyoshi, Y.; Stark, H., 2015: GraDeR: Membrane Protein Complex Preparation for Single-Particle Cryo-EM. *Structure*., **23**, 1769–1775.

He, M.; Sebaihia, M.; Lawley, T. D.; Stabler, R. A.; Dawson, L. F.; Martin, M. J.; Holt, K. E.; Seth-Smith, H. M. B.; Quail, M. A.; Rance, R.; Brooks, K.; Churcher, C.; Harris, D.; Bentley, S. D.; Burrows, C.; Clark, L.; Corton, C.; Murray, V.; Rose, G. et al., 2010: Evolutionary dynamics of Clostridium difficile over short and long time scales. *Proceedings of the National Academy of Sciences of the United States of America*., **107**, 7527–7532.

Herrera, M. G.; Noguera, M. E.; Sewell, K. E.; Agudelo Suárez, W. A.; Capece, L.; Klinke, S.; Santos, J., 2019: Structure of the Human ACP-ISD11 Heterodimer. *Biochemistry*., **58**, 4596–4609.

Hilbi, H.; Hoffmann, C.; Harrison, C. F., 2011: Legionella spp. outdoors: Colonization, communication and persistence. *Environmental Microbiology Reports*., **3**, 286–296.

Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R., 1989: Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*., **77**, 51–59.

Holm, L.; Laakso, L. M., 2016: Dali server update. *Nucleic acids research*., **44**, W351–W355.

Holm, L., 2019: Benchmarking fold detection by DaliLite v.5. *Bioinformatics*., **7**, 1–2.

Holmquist, M.; Clausen, I. G.; Patkar, S.; Svendsen, A.; Hult, K., 1995: Probing a functional role of Glu87 and Trp89 in the lid of Humicola lanuginosa lipase through transesterification reactions in organic solvent. *Journal of Protein Chemistry*., **14**, 217–224.

Hopfner, K. P.; Kopetzki, E.; Kreße, G. B.; Bode, W.; Huber, R.; Engh, R. A., 1998: New enzyme lineages by subdomain shuffling. *Proceedings of the National Academy of Sciences of the United States of America*., **95**, 9813–9818.

Isberg, R. R.; O'Connor, T. J.; Heidtman, M., 2009: The Legionella pneumophila replication vacuole: Making a cosy niche inside host cells. *Nature Reviews Microbiology*., **7**, 13–24.

Jackson, S.; Calos, M.; Myers, A.; Self, W. T., 2006: Analysis of Proline Reduction in the Nosocomial Pathogen Clostridium difficile �, **188**, 8487–8495.

Jeong, S. T.; Kim, H. K.; Kim, S. J.; Chi, S. W.; Pan, J. G.; Oh, T. K.; Ryu, S. E., 2002: Novel zinc-binding center and a temperature switch in the Bacillus stearothermophilus L1 lipase. *Journal of Biological Chemistry*., **277**, 17041–17047.

Jerabek-Willemsen, M.; André, T.; Wanner, R.; Roth, H. M.; Duhr, S.; Baaske, P.; Breitsprecher, D., 2014: MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*., **1077**, 101–113.

Jiang, Y.; Li, B. Q.; Zhang, Y.; Feng, Y. M.; Gao, Y. F.; Zhang, N.; Cai, Y. D., 2013: Prediction and Analysis of Post-Translational Pyruvoyl Residue Modification Sites from Internal Serines in Proteins. *PLoS ONE*., **8**, 13–15.

Jones, D. T., 1999: Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*., **292**, 195–202.

Kabisch, U. C.; Gräntzdörffer, A.; Schierhorn, A.; Rücknagel, K. P.; Andreesen, J. R.; Pich, A., 1999: Identification of D-proline reductase from Clostridium sticklandii as a selenoenzyme and indications for a catalytically active pyruvoyl group derived from a cysteine residue by cleavage of a proprotein. *Journal of Biological Chemistry*., **274**, 8445–8454.

Kabsch, W., 2010a: Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography*., **66**, 133–144.

Kabsch, W., 2010b: *XDS*. *Acta Crystallographica Section D Biological Crystallography*., **66**, 125–132.

Kamada, N.; Kim, Y. G.; Sham, H. P.; Vallance, B. A.; Puente, J. L.; Martens, E. C.; Núñez, G., 2012: Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science*., **336**, 1325–1329.

Karisson, S.; Burman, L. G.; Akerlund, T., 1999: Suppression of toxin production in Clostridium difficile VPI 10463 by amino acids, **145**, 1683–1693.

Kastner, B.; Fischer, N.; Golas, M. M.; Sander, B.; Dube, P.; Boehringer, D.; Hartmuth, K.;

Deckert, J.; Hauer, F.; Wolf, E.; Uchtenhagen, H.; Urlaub, H.; Herzog, F.; Peters, J. M.; Poerschke, D.; Lührmann, R.; Stark, H., 2008: GraFix: sample preparation for single-particle electron cryomicroscopy. *Nature methods*., **5**, 53–55.

Kawata, T.; Takeoka, A.; Takumi, K.; Masuda, K., 1984: Demonstration and preliminary characterization of a regular array in the cell wall of Clostridium difficile. *FEMS Microbiology Letters*., **24**, 323–328.

Kelly, C. P.; Pothoulakis, C.; LaMont, T., 1994: Clostridium difficile colitis. *Current Concepts*., **330**, 257–262.

Khan, F. I.; Lan, D.; Durrani, R.; Huan, W.; Zhao, Z.; Wang, Y., 2017: The lid domain in lipases: Structural and functional determinant of enzymatic properties. *Frontiers in Bioengineering and Biotechnology*., **5**, 1–13.

Knetsch, C. W.; Terveer, E. M.; Lauber, C.; Gorbalenya, A. E.; Harmanus, C.; Kuijper, E. J.; Corver, J.; van Leeuwen, H. C., 2012: Comparative analysis of an expanded Clostridium difficile reference strain collection reveals genetic diversity and evolution through six lineages. *Infection, Genetics and Evolution*., **12**, 1577–1585.

Knight, D. R.; Elliott, B.; Chang, B. J.; Perkins, T. T.; Riley, T. V., 2015: Diversity and evolution in the genome of clostridium difficile. *Clinical Microbiology Reviews*., **28**, 721–741.

Krebs, H. A., 1967: The redox state of nicotinamide adenine dinucleotide in the cytoplasm and mitochondria of rat liver. *Advances in Enzyme Regulation*., **5**, 409–434.

Krissinel, E.; Henrick, K., 2007: Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*., **372**, 774–797.

Kühlbrandt, W., 2014: The resolution revolution. *Science*., **343**, 1443–1444.

Kuhle, K.; Krausze, J.; Curth, U.; Rössle, M.; Heuner, K.; Lang, C.; Flieger, A., 2014: Oligomerization inhibits legionella pneumophila PlaB phospholipase A activity. *Journal of Biological Chemistry*., **289**, 18657–18666.

Lamontagne, F.; Labbé, A. C.; Haeck, O.; Lesur, O.; Lalancette, M.; Patino, C.; Leblanc, M.; Laverdière, M.; Pépin, J., 2007: Impact of emergency colectomy on survival of patients with fulminant Clostridium difficile colitis during an epidemic caused by a hypervirulent strain. *Annals of Surgery*., **245**, 267–272.

Lang, C.; Rastew, E.; Hermes, B.; Siegbrecht, E.; Ahrends, R.; Banerji, S.; Flieger, A., 2012: Zinc metalloproteinase ProA directly activates Legionella pneumophila PlaC glycerophospholipid:cholesterol acyltransferase. *Journal of Biological Chemistry*., **287**, 23464–23478.

Lang, C.; Hiller, M.; Flieger, A., 2017: Disulfide loop cleavage of Legionella pneumophila PlaA boosts lysophospholipase A activity. *Scientific Reports*., **7**, 1–14.

Leber, V.; Nans, A.; Singleton, M. R., 2018: Structural basis for assembly of the CBF3 kinetochore complex. *The EMBO Journal*., **37**, 269–281.

Louis-Jeune, C.; Andrade-Navarro, M.; Perez-Iratxeta, C., 2012: Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins: Structure, Function and Bioinformatics*., **80**, 374–381.

Lovitt, R. W.; Kell, D. B.; Morris, J. G., 1986: Proline reduction by Clostridium sporogenes is coupled to vectorial proton ejection. *FEMS Microbiology Letters*., **36**, 269–273.

Lubkowski, J.; Palm, G. J.; Gilliland, G. L.; Derst, C.; Röhm, K. H.; Wlodawer, A., 1996: Crystal structure and amino acid sequence of Wolinella succinogenes L-asparaginase. *European Journal of Biochemistry*., **241**, 201–207.

Lupp, C.; Robertson, M. L.; Wickham, M. E.; Sekirov, I.; Champion, O. L.; Gaynor, E. C.; Finlay, B. B., 2007: Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. *Cell Host and Microbe*., **2**, 119–129.

Mani, N.; Dupuy, B., 2001: Regulation of toxin synthesis in Clostridium difficile by an alternative RNA polymerase sigma factor. *Proceedings of the National Academy of Sciences of the United States of America*., **98**, 5844–5849.

Maret, W., 2012: New perspectives of zinc coordination environments in proteins. *Journal of Inorganic Biochemistry*., **111**, 110–116.

Markwick, P. R. L.; Malliavin, T.; Nilges, M., 2008: Structural biology by NMR: Structure, dynamics, and interactions. *PLoS Computational Biology*., **4**.

Martin-Verstraete, I.; Peltier, J.; Dupuy, B., 2016: The Regulatory Networks That Control Clostridium difficile Toxin Synthesis. *Toxins*., **8**, 1–24.

Matamouros, S.; England, P.; Dupuy, B., 2007: Clostridium difficile toxin expression is inhibited by the novel regulator TcdC. *Molecular Microbiology*., **64**, 1274–1288.

Matsumara, H.; Yamamoto, T.; Leow, T. C.; Mori, T.; Salleh, A. B.; Basri, M.; Inoue, T.; Kai, Y.; Rahman, R. N. Z. R. A., 2008: Novel cation-pi interaction reveales by crystal structure of thermoalkalophilic lipase. *Proteins: Structure, Function and Bioinformatics*., **70**, 592–598.

McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., 2007: Phaser crystallographic software. *Journal of Applied Crystallography*., **40**, 658–674.

McDonald, L. C.; Gerding, D. N.; Johnson, S.; Bakken, J. S.; Carroll, K. C.; Coffin, S. E.; Dubberke, E. R.; Garey, K. W.; Gould, C. V.; Kelly, C.; Loo, V.; Shaklee Sammons, J.; Sandora, T. J.; Wilcox, M. H., 2018: Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clinical Infectious Diseases*., **66**, e1–e48.

Meiler, J.; Baker, D., 2003: Rapid protein fold determination using unassigned NMR data. *Proceedings of the National Academy of Sciences of the United States of America*., **100**, 15404–15409.

Merk, A.; Bartesaghi, A.; Banerjee, S.; Falconieri, V.; Rao, P.; Davis, M. I.; Pragani, R.; Boxer, M. B.; Earl, L. A.; Milne, J. L. S.; Subramaniam, S., 2016: Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell*., **165**, 1698–1707.

Mindrebo, J.; Nartey, C.; Seto, Y.; Burkart, M.; Noel, J., 2016: Unveilling the functional diversity of the Alpha-Beta hydrolase fold in plants. *Current opinion in structural biology*., **41**, 233–246.

Monot, M.; Boursaux-Eude, C.; Thibonnier, M.; Vallenet, D.; Moszer, I.; Medigue, C.; Martin-Verstraete, I.; Dupuy, B., 2011: Reannotation of the genome sequence of Clostridium difficile strain 630. *Journal of Medical Microbiology*., **60**, 1193–1199.

Moon, C. P.; Fleming, K. G., 2011: *Using tryptophan fluorescence to measure the stability of membrane proteins folded in liposomes*. *Methods in Enzymology*. 1st edn. Elsevier Inc., Vol. 492.

Moore, R. J.; Lacey, J. A., 2019: Genomics of the Pathogenic Clostridia. *Microbiology Spectrum*., **7**, 1–17.

Mylonas, E.; Svergun, D. I., 2007: Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal of Applied Crystallography*., **40**, 245–249.

Newton, H. J.; Ang, D. K. Y.; Van Driel, I. R.; Hartland, E. L., 2010: Molecular pathogenesis of infections caused by Legionella pneumophila. *Clinical Microbiology Reviews*., **23**, 274–298.

Novick, P.; Zerial, M., 1997: The diversity of Rab proteins in vesicle transport. *Current Opinion in Cell Biology*., **9**, 496–504.

Ong, G. K. B.; Reidy, T. J.; Huk, M. D.; Lane, F. R., 2017: Clostridium difficile colitis: A clinical review. *American Journal of Surgery*., **213**, 565–571.

Ovchinnikov, S.; Kamisetty, H.; Baker, D., 2014: Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife*., **2014**, 1–21.

Pace, N. J.; Weerapana, E., 2014: Zinc-binding cysteines: diverse functions and structural motifs. *Biomolecules*., **4**, 419–434.

Pape, T.; Schneider, T. R., 2004: *HKL2MAP* : a graphical user interface for macromolecular phasing with *SHELX* programs. *Journal of Applied Crystallography*., **37**, 843–844.

Patel, B. A.; Gospodarek, A.; Larkin, M.; Kenrick, S. A.; Haverick, M. A.; Tugcu, N.; Brower, M. A.; Richardson, D. D., 2018: Multi-angle light scattering as a process analytical technology measuring real-time molecular weight for downstream process control. *mAbs*., **10**, 945–950.

Pauw, B. R., 2013: Everything SAXS: Small-angle scattering pattern collection and correction. *Journal of Physics: Condensed Matter*., **25**, 1–24.

Perelle, S.; Gibert, M.; Bourlioux, P.; Corthier, G.; Popoff, M. R., 1997: Production of a complete binary toxin (actin-specific ADP- ribosyltransferase) by Clostridium difficile CD196. *Infection and Immunity*., **65**, 1402–1407.

Perutz, M.; Rossmann, M.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T., 1960: Structure of haemoglobin. *Nature*., **185**, 416–422.

Peterfreund, G. L.; Vandivier, L. E.; Sinha, R.; Marozsan, A. J.; Olson, W. C.; Zhu, J.; Bushman, F. D., 2012: Succession in the Gut Microbiome following Antibiotic and Antibody Therapies for Clostridium difficile. *PLoS ONE*., **7**.

Poehlein, A.; Alghaithi, H. S.; Chandran, L.; Chibani, C. M.; Davydova, E.; Dhamotharan, K.; Ge, W.; Gutierrez-Gutierrez, D. A.; Jagirdar, A.; Khonsari, B.; Prakash R Nair, K. P.; Daniel, R.; Poehlein, C. A.; Kppr, N., 2014: First Insights into the Genome of the Amino Acid-Metabolizing Bacterium Clostridium litorale DSM 5388. *Genome Announcements*., **2**, e00754-14.

Poulsen, C.; Holton, S.; Geerlof, A.; Wilmanns, M.; Song, Y. H., 2010: Stoichiometric protein complex formation and over-expression using the prokaryotic native operon structure. *FEBS Letters*., **584**, 669–674.

Prasadarao, N. V.; Wass, C. A.; Weiser, J. N.; Stins, M. F.; Huang, S. H. E.; Kim, K. S., 1996: Outer membrane protein A of Escherichia coli contributes to invasion of brain microvascular endothelial cells. *Infection and Immunity*., **64**, 146–153.

Purslow, J. A.; Khatiwada, B.; Bayro, M. J.; Venditti, V., 2020: NMR Methods for Structural Characterization of Protein-Protein Complexes. *Frontiers in Molecular Biosciences*., **7**, 1–8.

Rauwerdink, A.; Kazlauskas, R., 2015: How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: the Serine-Histidine-Aspartate Catalytic Triad of a/b-Hydrolase Fold Enzymes. *ACS Catalysis*., **5**, 6153–6176.

Reeves, A. E.; Theriot, C. M.; Bergin, I. L.; Huffnagle, G. B.; Schloss, P. D.; Young, V. B., 2011: The interplay between microbiome dynamics and pathogen dynamics in a murine model of Clostridium difficile infection. *Gut Microbes*., **2**, 145–158.

Remmert, M.; Biegert, A.; Hauser, A.; Söding, J., 2012: HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods*., **9**, 173–175.

Robert Koch-Institut, 2019: *Infektionsepidemiologisches Jahrbuch meldepflichtiger Krankheiten für 2018*. (Askar, M., Ed.). Robert Koch-Institut, Berlin.

Rodriguez-Palacios, A.; Stämpfli, H. R.; Duffield, T.; Peregrine, A. S.; Trotz-Williams, L. A.; Arroyo, L. G.; Brazier, J. S.; Weese, J. S., 2006: Clostridium difficile PCR ribotypes in calves, Canada. *Emerging Infectious Diseases*., **12**, 1730–1736.

Rodriguez-Palacios, A.; Stämpfli, H. R.; Stalker, M.; Duffield, T.; Weese, J. S., 2007: Natural and experimental infection of neonatal calves with Clostridium difficile. *Veterinary Microbiology*., **124**, 166–172.

Roy, C. R.; Isberg, R. R., 1997: Topology of Legionella pneumophila DotA: An inner membrane protein required for replication in macrophages. *Infection and Immunity*., **65**, 571–578.

Rozbeský, D.; Rosůlek, M.; Kukačka, Z.; Chmelík, J.; Man, P.; Novák, P., 2018: Impact of Chemical Cross-Linking on Protein Structure and Function. *Analytical Chemistry*., **90**, 1104–1113.

Rupp, B., 2009: *Biomolecular Crystallography: Principles, Practice, and Application to Structural*

Biology. Taylor & Francis Ltd.

Ruska, E., 1987: The Development of the Electron Microscope and of Electron Microscopy. *Angewandte Chemie (International ed. in English)*., **26**, 595–706.

Sanches, M.; Barbosa, J. A. R. G.; De Oliveira, R. T.; Neto, J. A.; Polikarpov, I., 2003: Structural comparison of Escherichia coli L-asparaginase in two monoclinic space groups. *Acta Crystallographica - Section D Biological Crystallography*., **59**, 416–422.

Scheres, S. H. W., 2016: *Processing of Structurally Heterogeneous Cryo-EM Data in RELION*. *Methods in Enzymology*. 1st edn. Elsevier Inc., Vol. 579.

Schleberger, C.; Sachelaru, P.; Brandsch, R.; Schulz, G. E., 2007: Structure and Action of a C-C Bond Cleaving α/β-Hydrolase Involved in Nicotine Degration. *Journal of Molecular Biology*., **367**, 409–418.

Schmidt, C.; Urlaub, H., 2017: Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies. *Current Opinion in Structural Biology*., **46**, 157–168.

Schmiel, D. H.; Miller, V. L., 1999: Bacterial phospholipases and pathogenesis. *Microbes and Infection*., **1**, 1103–1112.

Schmitzberger, F.; Kilkenny, M. L.; Lobley, C. M. C.; Webb, M. E.; Vinkovic, M.; Matak-Vinkovic, D.; Witty, M.; Chirgadze, D. Y.; Smith, A. G.; Abel, C.; Blundell, T. L., 2003: Structural constraints on protein self-processing in L-aspartate-α -decarboxylase. *EMBO Journal*., **22**, 6193–6204.

Schrag, J. D.; Li, Y.; Cygler, M.; Lang, D.; Burgdorf, T.; Hecht, H. J.; Schmid, R.; Schomburg, D.; Rydel, T. J.; Oliver, J. D.; Strickland, L. C.; Dunaway, C. M.; Larson, S. B.; Day, J.; McPherson, A., 1997: The open conformation of a Pseudomonas lipase. *Structure*., **5**, 187–202.

Schunder, E.; Adam, P.; Higa, F.; Remer, K. A.; Lorenz, U.; Bender, J.; Schulz, T.; Flieger, A.; Steinert, M.; Heuner, K., 2010: Phospholipase PlaB is a new virulence factor of Legionella pneumophila. *International Journal of Medical Microbiology*., **300**, 313–323.

Sebaihia, M.; Wren, B. W.; Mullany, P.; Fairweather, N. F.; Minton, N.; Stabler, R.; Thomson, N. R.; Roberts, A. P.; Cerdeño-Tárraga, A. M.; Wang, H.; Holden, M. T. G.; Wright, A.; Churcher, C.; Quail, M. A.; Baker, S.; Bason, N.; Brooks, K.; Chillingworth, T.; Cronin, A. et al., 2006: The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nature Genetics*., **38**, 779–786.

Secundo, F.; Carrea, G.; Tarabiono, C.; Brocca, S.; Lotti, M., 2004: Activity and Enantioselectivity of Wildtype and Lid Mutated Candida rugosa Lipase Isoform 1 in Organic Solvents. *Biotechnology and Bioengineering*., **86**, 236–240.

Selleck, W.; Tan, S., 2008: Recombinant protein complex expression in E. coli. *Current Protocols in Protein Science*., 1–25.

Seto, B.; Stadtman, T. C., 1976: Purification and properties of proline reductase from Clostridium sticklandii. *Journal of Biological Chemistry*., **251**, 2435–2439.

Seto, B., 1978: A Pyruvate-containing Peptide of Proline Reductase in Clostridium sticklandii, **253**, 4525–4529.

Sharp, K. A., 1996: Electrostatic interactions in hirudin-thrombin binding. *Biophysical Chemistry*., **61**, 37–49.

Shen, A., 2012: Clostridium difficile toxins: Mediators of inflammation. *Journal of Innate Immunity*., **4**, 149–158.

Sirard, S.; Valiquette, L.; Fortier, L. C., 2011: Lack of association between clinical outcome of Clostridium difficile infections, strain type, and virulence-associated phenotypes. *Journal of Clinical Microbiology*., **49**, 4040–4046.

Smits, W. K.; Lyras, D.; Lacy, B.; Wilcox, M. H.; Kuijper, E. J., 2016: Clostridium difficile infection. *Nature Reviews Disease Primers*., **2**.

Song, W. J.; Sontz, P. A.; Ambroggio, X. I.; Tezcan, F. A., 2014: Metals in Protein–Protein Interfaces. *Annual Review of Biophysics*., **43**, 409–431.

Stark, H., 2010: GraFix: Stabilization of fragile macromolecular complexes for single particle Cryo-EM. *Methods in Enzymology*., **481**, 109–126.

Steele, J.; Parry, N.; Tzipori, S., 2013: The roles of toxin A and toxin B in Clostridium difficile infection: Insights from the gnotobiotic piglet model. *Gut Microbes*., **5**, 53–57.

Stickland, L. H., 1934: Studies of the metabolism of the strict anaerobes (genus Clostridium). The chemical reactions by which Cl. sporogenes obtains its energy. *Biochemical Journal*., **28**, 1746–1759.

Stott, K.; Blackburn, J. M.; Butler, P. J. G.; Perutz, M., 1995: Incorporation of glutamine repeats makes protein oligomerize: Implications for neurodegenerative diseases. *Proceedings of the*

National Academy of Sciences of the United States of America., **92**, 6509–6513.

Studier, F. W.; Moffatt, B. A., 1986: Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*., **189**, 113–130.

Sun, C.; Gennis, R. B., 2019: Single-particle cryo-EM studies of transmembrane proteins in SMA copolymer nanodiscs. *Chemistry and Physics of Lipids*., **221**, 114–119.

Suplatov, D. A.; Besenmatter, W.; Švedas, V. K.; Svendsen, A., 2012: Bioinformatic analysis of alpha/beta-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities. *Protein Engineering, Design and Selection*., **25**, 689–697.

Surawicz, C. M.; Brandt, L. J.; Binion, D. G.; Ananthakrishnan, A. N.; Curry, S. R.; Gilligan, P. H.; McFarland, L. V.; Mellow, M.; Zuckerbraun, B. S., 2013: Guidelines for diagnosis, treatment, and prevention of clostridium difficile infections. *American Journal of Gastroenterology*., **108**, 478–498.

Svergun, D. I.; Koch, M. H. J., 2003: Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*., **66**, 1735–1782.

Swanson, M. S.; Hammer, B. K., 2000: *Legionella Pneumophila* Pathogenesis: A Fateful Journey from Amoebae to Macrophages. *Annual Review of Microbiology*., **54**, 567–613.

Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Zwart, P. H.; Hung, L. W.; Read, R. J.; Adams, P. D., 2007: Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D: Biological Crystallography*., **64**, 61–69.

Tiesinga, J. J. W.; Pouderoyen, G. van; Nardini, M.; Ransac, S.; Dijkstra, B. W., 2007: Structural Basis of Phospholipase Activity of Staphylococcus hyicus lipase. *Journal of Molecular Biology*., **371**, 447–456.

Tolbert, W. D.; Graham, D. E.; White, R. H.; Ealick, S. E., 2003: Pyruvoyl-dependent arginine decarboxylase from Methanococcus jannaschii: Crystal structures of the self-cleaved and S53A proenzyme forms. *Structure*., **11**, 285–294.

Tremblay, P. L.; Zhang, T.; Dar, S.; Leang, C.; Lovley, D. R., 2012: The Rnf Complex of Clostridium ljungdahlii Is a Proton-Translocating Ferredoxin:NADɦ Oxidoreductase Essential for Autotrophic Growth. *mBio*., **4**, 1–8.

Trip, H.; Mulder, N. L.; Rattray, F. P.; Lolkema, J. S., 2011: HdcB, a novel enzyme catalysing maturation of pyruvoyl-dependent histidine decarboxylase. *Molecular Microbiology*., **79**, 861–871.

Truchan, H. K.; Christman, H. D.; White, R. C.; Rutledge, N. S.; Cianciotto, N. P., 2017: Type II Secretion Substrates of Legionella pneumophila Translocate Out of the Pathogen-Occupied Vacuole via a Semipermeable Membrane. *mBio*., **8**, 1–18.

Tyndall, J. D. A.; Sinchaikul, S.; Fothergill-Gilmore, L. A.; Taylor, P.; Walkinshaw, M. D., 2002: Crystal structure of a thermostable lipase from Bacillus stearothermophilus P1. *Journal of Molecular Biology*., **323**, 859–869.

van Eijk, E.; Anvar, S. Y.; Browne, H. P.; Leung, W. Y.; Frank, J.; Schmitz, A. M.; Roberts, A. P.; Smits, W. K., 2015: Complete genome sequence of the Clostridium difficile laboratory strain 630Δerm reveals differences from strain 630, including translocation of the mobile element CTn5. *BMC genomics*., **16**, 31.

Voth, D. E.; Ballard, J. D., 2005: Clostridium diffcile Toxins: Mechanism of Action and Role in Disease. *Clinical Microbiology Reviews*., **18**, 247–263.

Vranckx, L.; De Buck, E.; Anné, J.; Lammertyn, E., 2007: Legionella pneumophila exhibits plasminogen activator activity. *Microbiology*., **153**, 3757–3765.

Wang, F.; Yu, Z.; Betegon, M.; Campbell, M. G.; Aksel, T.; Zhao, J.; Li, S.; Douglas, S. M.; Cheng, Y.; Agard, D. A., 2020: Amino and PEG-amino graphene oxide grids enrich and protect samples for high-resolution single particle cryo-electron microscopy. *Journal of Structural Biology*., **209**, 107437.

Warny, M.; Pepin, J.; Fang, A.; Killgore, G.; Thompson, A.; Brazier, J.; Frost, E.; McDonald, L. C., 2005: Toxin production by an emerging strain of Clostridium difficile associated with outbreaks of severe disease in North America and Europe. *Lancet*., **30**, 1079–1084.

Weese, J. S., 2010: Clostridium difficile in food - innocent bystander or serious threat? *Clinical Microbiology and Infection*., **16**, 3–10.

Widmer, H.; Billeter, M.; Wüthrich, K., 1989: Three-dimensional structure of the neurotoxin ATX Ia from Anemonia sulcata in aqueous solution determined by nuclear magnetic resonance spectroscopy. *Proteins: Structure, Function, and Bioinformatics*., **6**, 357–371.

Wienken, C. J.; Baaske, P.; Rothbauer, U.; Braun, D.; Duhr, S., 2010: Protein-binding assays in biological liquids using microscale thermophoresis. *Nature communications*., **1**, 100.

Williamson, M. P.; Havel, T. F., 1985: Solution conformation of Proteinase Inhibitor IIa from Bull Seminal plasma by H NMR and distance geometry. *J. Mol. Biol.*, **182**, 295–315.

Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., 2011: Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*., **67**, 235–242.

Winn, W. C.; Myerowitz, R. L., 1981: The pathology of the legionella pneumonias. *Human Pathology*., **12**, 401–422.

Wishart, D. S.; Case, D. A., 2002: Use of chemical shifts in macromolecular structure determination. *Methods in Enzymology*., **338**, 3–34.

Wyatt, P. J., 1998: Submicrometer particle sizing by multiangle light scattering following fractionation. *Journal of Colloid and Interface Science*., **197**, 9–20.

Xu, Y.; Zheng, Y.; Fan, J. S.; Yang, D., 2006: A new strategy for structure determination of large proteins in solution without deuteration. *Nature Methods*., **3**, 931–937.

Yaeger, M. J.; Kinyon, J. M.; Songer, J. G., 2007: A prospective, case control study evaluating the association between Clostridium difficile toxins in the colon of neonatal swine and gross and microscopic lesions. *Journal of Veterinary Diagnostic Investigation*., **19**, 52–59.

Zhao, Z.; Hou, S.; Lan, D.; Wang, X.; Liu, J.; Khan, F. I.; Wang, Y., 2017: Crystal structure of a lipase from Streptomyces sp. strain W007 – implications for thermostability and regiospecificity. *FEBS Journal*., **284**, 3506–3519.

Zi Tan, Y.; Baldwin, P. R.; Davis, J. H.; Williamson, J. R.; Potter, C. S.; Carragher, B.; Lyumkis, D., 2017: Addressing preferred specimen orientation in single-particle cryo-EMthrough tilting. *Nature Methods*., **14**, 793–796.

9 Supplementary

Table S1: Complete list of primers used in the projects presented in this thesis. Primer short names are provided for assignment to addressed gene, purpose, direction and restriction enzyme (RE) used and melting temperature Tm.

Primer name	Gene	Start/End Mutation	↔	RE	Sequence	Tm [°C]
MD001	PrdA	M1	fw	NdeI	GGAATTCCATATGTCAATAACTTTAGAAACAGCTCAAG	66
MD002	PrdA	I422	fw	NdeI	GGAATTCCATATGATAGGACCTGCATCAAAAGAAATG	66
MD003	PrdB	M1	fw	NdeI	GGAATTCCATATGAGCCTTACAACAGTACAAGGAC	66
MD007	PrdA	T420	rv	BamHI	CGGGATCCttaGTTAATACCATTACAGCTCCATC	64
MD008	PrdA	N626	rv	BamHI	CGGGATCCTTAGTTTTCGTCTTTTTCGTATATTTC	68
MD009	PrdB	V241	rv	BamHI	CGGGATCCTTAAACGTGAGCTTTATATTCGTATG	68
MD014	PrdB	U151C	rv		CAAGTACCGCAGCCAGCTGTC	68
MD016	PrdB	U151C	fw		GACAGCTGGCTGCGGTACTTG	68
MD020	PrdA	M1	fw	NotI	AAGGAAAAAAGCGGCCGCTCAATAACTTTAGAAACAGCTCAAG	
MD021	PrdA	T420	rv	KpnI	GGGGTACCttaGTTAATACCATTACAGCTCCATC	
MD022	PrdA	P164	fw	NdeI	tttcagggccatatgGGAATTCCATATGCCAAAAGTTGAAGAAAAGAAAC	58
MD023	PrdA	T420	rv	XhoI	ttagcagccggatccctcgagttaAGTTAATGCATGGATACCACC	60
MD024	PrdA	I422	fw	NdeI	tttcagggccatatgGGAATTCCATATGATAGGACCTGCATCAAAAGAA	60
MD025	PrdA	V573	rv	XhoI	ttagcagccggatccctcgagttaTACACCTTCTATAGCATTTTTAAG	60
MD026	PrdA	A149	fw	NdeI	GGAATTCCATATGGCAAACGTTGCTGCTCCAGTG	68
MD027	PrdA	A158	fw	NdeI	GGAATTCCATATGGCAGCAGCAGAAGTAGCTCC	64
MD028	PrdB	M1	fw	NotI	ATAAGAATGCGGCCGCATGAGCCTTACAACAGTACAAGGAC	
MD029	PrdB	V241	rv	KpnI	GGGGTACCTTAAACGTGAGCTTTATATTCGTATG	
MD030	PrdB	V241	rv	KpnI	GGGGTACCAACGTGAGCTTTATATTCGTATGGTA	
MD031	PrdA	S174	fw	NotI	ATAAGAATGCGGCCGCATGAGCTTAACTAAAAAACACTTTAAAATAG	68

MD032	PrdA	G198	fw	NotI	ATAAGAATGCGGCCGCATGGGAACTACTTTATACATCAGAAAC	64
MD033	PrdA	N626	rv	KpnI	GGGGTACCTTAGTTTTCGTCTTTTTCGTATATTTC	
MD038	PrdB	V241	rv	BamHI	CGGGATCCTTAAACGTGAGCTTTATATTCGTATG	68
MD039	PrdB	I223	rv	BamHI	CGGGATCCTTATATTGCAACTAATTGCTCTAATGTAT	66
MD040	PrdB	S11	fw	NotI	AAGGAAAAAAGCGGCCGCATGTCTGAAATATTCGTTCCAATAACAC	66
MD041	PrdB	V22	fw	NotI	AAGGAAAAAAGCGGCCGCATGGTTTGGACTCCTGTAACTAAAG	64
MD042	PrdA	E622	rv	BamHI	CGGGATCCTTATTCGTATATTTCTCTTCTCTTC	64
MD043	PrdA	E161A	fw		GTGCAGCAGCAGCAGTAGCTCCAAAAG	86
MD044	PrdA	E161A	rv		CTTTTGGAGCTACTGCTGCTGCTGCAC	86
MD045	PrdA	E355A	fw		GAAGAAGTAATAGCGCAATACAGAAGAG	78
MD046	PrdA	E355A	rv		CTCTTCTGTATTGCGCTATTACTTCTTC	78
MD047	PrdB	M1	fw	NotI	AGCCTTACAACAGTACAAGGAC	
MD66	PrdA	E81Stop	fw		CAGATAACGTAGAATGATACAAAGCAG	
MD67	PrdA	E81Stop	rv		CTGCTTTGTATCATTCTACGTTATCTG	
MD68	PrdA	G117Stop	fw		GAGTTGCAGGTTAAACAGTTAAAATAC	
MD69	PrdA	G117Stop	rv		GTATTTTAACTGTTTAACCTGCAACTC	
MD70	PrdA	G125Stop	fw		AAATACACATAGGTTAAGGTAAGAACATAA	
MD71	PrdA	G125Stop	rv		TTATGTTCTTACCTTAACCTATGTGTATTT	
MD72	PrdA	G140Stop	fw		ATCAGTAGCTTGATAAGCAGGAGTTG	
MD73	PrdA	G140Stop	rv		CAACTCCTGCTTATCAAGCTACTGAT	
MD74	PrdA	V166Stop	fw		GCTCCAAAAGTTTAAGAAAAGAAACTTTT	
MD75	PrdA	V166Stop	rv		AAAAGTTTCTTTTCTTAAACTTTTGGAGC	
MD76	PrdA	D189Stop	fw		GAATTTGCTGATTAAACTAAAATAGAAG	
MD77	PrdA	D189Stop	rv		CTTCTATTTTAGTTTAATCAGCAAATTC	
MD78	PrdB	E29	fw	NotI	ATAAGAATGCGGCCGCATGGAATTAAAAGATATGACTATAGC	
MD79	PrdB	M33	fw	NotI	ATAAGAATGCGGCCGCATGACTATAGCTTTAGCTACAGCAG	
MD80	PrdB	A66	fw	NotI	ATAAGAATGCGGCCGCATGGCTACAGTTGATGAAATGATGG	

Table S3: Percent identity matrix for a multiple sequence alignment of PlaB *Legionella pneumophila str Corby* generated by Clustal2.1. The identity for different *Legionella* does not fall below 79%. Distant homologues can be found with an identity down to 23% in other lung pathogens like *Pseudomonas aeruginosa*.

		1 CAD92451.1 phospholipase B [Legionella pneumophila str. Corby]	2 WP 014841714.1 hypothetical protein [Legionella pneumophila]	3 WP 058515395.1 hypothetical protein [Legionella santicrucis]	4 WP 058483282 1 hypothetical protein [Legionella spiritensis]	5 WP 045094542.1 hypothetical protein [Legionella fallonii]	6 WP 058463789.1 hypothetical protein [Legionella cincinnatiensis]	7 WP 019234277 1 hypothetical protein [Legionella anisa]	8 WP 143869021.1 phospholipase [Legionella israelensis]	9 WP 131795480.1 phospholipase [Fluoribacter gormanii]	10 WP 018577989.1 hypothetical protein [Legionella shakespearei]	11 WP 108292309.1 phospholipase [Legionella taurinensis]	12 WP 058532019.1 hypothetical protein [Legionella rubrilucens]	13 WP 011768684 1 hypothetical protein [Psychromonas ingrahamii]	14 WP 027722894 1 hypothetical protein [Desulfovibrio zosterae]	15 WP 012153585 1 phospholipase [Shewanella pealeana]	16 WP 053813721 1 hypothetical protein [Pseudomonas aeruginosa]	17 WP 012676954 1 hypothetical protein [Persephonella marina]	18 WP 007152182 1 hypothetical protein [Marinobacter algicola]
1	CAD92451.1 phospholipase B [Legionella pneumophila str. Corby]	100.00	98.10	88.61	88.82	88.82	87.76	86.92	86.92	85.23	83.76	79.75	79.11	45.98	44.93	44.57	25.85	24.47	23.88
2	WP 014841714.1 hypothetical protein [Legionella pneumophila]	98.10	100.00	88.19	88.82	88.61	87.34	86.50	86.50	85.23	83.76	79.32	78.90	45.76	44.71	43.65	25.37	24.24	23.40
3	WP 058515395.1 hypothetical protein [Legionella santicrucis]	88.61	88.19	100.00	88.19	91.98	95.99	93.46	90.72	85.44	85.86	81.65	81.43	45.31	44.71	43.42	43.76	25.18	25.53
4	WP 058483282 1 hypothetical protein [Legionella spiritensis]	88.82	88.82	88.19	100.00	89.87	87.97	86.29	88.40	86.71	85.02	79.32	79.11	46.43	45.59	44.34	25.37	25.88	24.82
5	WP 045094542.1 hypothetical protein [Legionella fallonii]	88.82	88.61	91.98	89.87	100.00	92.19	90.08	92.19	86.08	87.34	81.65	81.65	45.98	46.04	43.65	43.76	25.18	24.82
6	WP 058463789.1 hypothetical protein [Legionella cincinnatiensis]	87.76	87.34	95.99	87.97	92.19	100.00	93.04	90.93	85.23	85.23	82.28	81.65	45.31	45.81	43.42	25.85	24.47	25.30
7	WP 019234277 1 hypothetical protein [Legionella anisa]	86.92	86.50	93.46	86.29	90.08	93.04	100.00	89.24	85.02	84.60	81.86	81.43	45.31	44.05	43.88	43.82	24.00	43.64
8	WP 143869021.1 phospholipase [Legionella israelensis]	86.92	86.50	90.72	88.40	92.19	90.93	89.24	100.00	85.86	85.86	82.07	81.86	45.76	45.15	43.88	25.85	24.71	25.30
9	WP 131795480.1 phospholipase [Fluoribacter gormanii]	85.23	85.23	85.44	86.71	86.08	85.23	85.02	85.86	100.00	84.39	79.11	78.48	44.87	45.37	42.73	25.61	24.71	43.79
10	WP 018577989.1 hypothetical protein [Legionella shakespearei]	83.76	83.76	85.86	85.02	87.34	85.23	84.60	85.86	84.39	100.00	81.43	82.07	46.43	44.71	44.34	25.61	25.41	24.82
11	WP 108292309.1 phospholipase [Legionella taurinensis]	79.75	79.32	81.65	79.32	81.65	82.28	81.86	82.07	79.11	81.43	100.00	95.57	44.87	45.37	42.73	24.88	25.18	24.82
12	WP 058532019.1 hypothetical protein [Legionella rubrilucens]	79.11	78.90	81.43	79.11	81.65	81.65	81.43	81.86	78.48	82.07	95.57	100.00	45.98	44.71	43.42	43.82	25.41	24.35
13	WP 011768684 1 hypothetical protein [Psychromonas ingrahamii]	45.98	45.76	45.31	46.43	45.98	45.31	45.31	45.76	44.87	46.43	44.87	45.98	100.00	46.75	44.19	25.75	23.84	25.43
14	WP 027722894 1 hypothetical protein [Desulfovibrio zosterae]	44.93	44.71	44.71	45.59	46.04	45.81	44.05	45.15	45.37	44.71	45.37	44.71	46.75	100.00	47.97	26.55	24.76	25.55
15	WP 012153585 1 phospholipase [Shewanella pealeana]	44.57	43.65	43.42	44.34	43.65	43.42	43.88	43.88	42.73	44.34	42.73	43.42	44.19	47.97	100.00	27.78	25.93	26.00
16	WP 053813721 1 hypothetical protein [Pseudomonas aeruginosa]	25.85	25.37	43.76	25.37	43.76	25.85	43.82	25.85	25.61	25.61	24.88	43.82	25.75	26.55	27.78	100.00	28.87	45.50
17	WP 012676954 1 hypothetical protein [Persephonella marina]	24.47	24.24	25.18	25.88	25.18	24.47	24.00	24.71	24.71	25.41	25.18	25.41	23.84	24.76	25.93	28.87	100.00	34.78
18	WP 007152182 1 hypothetical protein [Marinobacter algicola]	23.88	23.40	25.53	24.82	24.82	25.30	43.64	25.30	43.79	24.82	24.82	24.35	25.43	25.55	26.00	45.50	34.78	100.00

10 Danksagung

Bei der Anfertigung dieser Arbeit habe ich unermesslich viel Unterstützung von sehr vielen verschiedenen Seiten bekommen. Ohne diese wäre ich nicht in der Lage gewesen, dieses Projekt zu stemmen und ich möchte mich bei all jenen bedanken. Es ist kaum möglich, allen Beteiligten gerecht zu werden.

Zunächst möchte ich meinem Doktorvater, Prof. Dr. Wulf Blankenfeldt, danken. Ich durfte bei Dir auf einem Thema promovieren, das mich gleichermaßen gefordert wie motiviert hat. Dabei hast Du mich unterstützt und beraten, mir aber gleichzeitig das Vertrauen zu den wissenschaftlichen Freiräumen entgegengebracht, das uns dorthin geführt hat, wo wir nun stehen. Dies habe ich jederzeit zu schätzen gewusst.

Herrn Prof. Dr. Dieter Jahn danke ich für die Übernahme der Zweitkorrektur. Ebenso gilt mein Dank Herrn Prof. Dr. Michael Hust für die Übernahme des Prüfungsvorsitzes.

Durch Frau Prof. Dr. Antje Flieger hatte ich die Chance, diese Arbeit um einen spannenden Themenbereich zu erweitern. Die Kooperation mit dem Robert Koch Institut und allen Beteiligten hat mir enorm viel Freude bereitet.

Meine Thesis Committees wurden durch die Herren Prof. Dr. Ralf Mendel, Prof. Dr. Andreas Pich, Prof. Dr. Dietmar Schomburg und Junior-Prof. Dr. Alexander Marchanka zu fruchtvollen Diskussionsrunden. Entscheidend für eine umfassende Arbeit sind viele verschiedene Blickwinkel.

Die integrative Beleuchtung der Projekte war nur mit der Hilfe von vielen helfenden Händen möglich. Angefangen bei der Bereitstellung von Mikroorganismen und Gedankenaustausch, bis hin zur Aufnahme von elektronenmikroskopischen Abbildungen, haben mich die Kooperationspartnerschaften mit Frau Dr. Meina Neumann-Schaal, Herrn Dr. Mathias Müsken und Frau Dr. Deniz Ugurlar begleitet.

Mein Interesse, sowie die frühen Erfolgserlebnisse in der Strukturbiologie verdanke ich dem Engagement von Herrn Dr. Andrea Scrima, Frau Terese Bergfors und Frau Prof. Dr. Maria Selmer, die mit mir die ersten strukturbiologischen Schritte gegangen sind. Ihr habt maßgeblich dazu beigetragen, dass ich diesen Weg einschlage.

Die alltäglichen Herausforderungen der Promotion wurden so viel bezwingbarer gemacht durch die Freundschaften, die während dieser Zeit geschlossen wurden. Danke lieber Dr. Stefan Schmelz, Dr. Peer Lukat, Dr. Jan Pippel und Dr. Konrad Büssow. Eure Hilfsbereitschaft war immer grenzenlos und sucht ihresgleichen.

Nach dem Motto: „Geteiltes Leid ist halbes Leid“, waren meine Kollegen immer da, um den Grill anzufeuern, Whiskey zu verkosten, zu feiern, zu tanzen und zu lachen. Mit euch habe ich viele schöne Momente erlebt, die ich nie vergessen werde. Danke Dr. Nina Schwemmlein, Dr. Florian Witzgall, Dr. Monika Popp, Dr. Thomas Klünemann, Dr. Björn Bulitta, Steffi Henke, ihr seid die besten!

Manchmal braucht man einen anderen Blickwinkel um klarer zu sehen. Wo könnte man besser andere Blickwinkel finden, wenn nicht im internationalen Austausch oder bei der Betreuungsarbeit? Dr. Archna Archna, Dr. Pooja Sadana, Dr. Kumar Siddarth „Sid“ Singh, Falain Fongue Zeukang, Sebastian Henz, ihr seid nicht nur Freunde und gute Kollegen, sondern auch wertvolle Augenöffner! Lieber Christian Behlendorf, ich drücke Dir die Daumen und wünsche Dir, dass Du ebenso viel Freude an der Weiterführung dieses Projektes findest wie ich.

Ein ebenfalls nicht unerheblicher Anteil an der Gestaltung meiner Ausbildung tragen die Graduiertenschulen PROCOMPAS und GS-FIRE, mit deren Mitteln ich mich in Europa und Amerika wissenschaftlich weiterbilden konnte. Der Kollegiale und freundschaftliche Austausch innerhalb der Graduiertenschulen waren immer spannend.

Und dann gibt es noch eine endlos lange Liste von helfenden Händen, vertrauensvollen Seelsorgern und gewissenhaften Organisatoren. Euer Beitrag ist häufig unsichtbar für andere. Und dennoch bildet ihr die Basis und den Klebstoff für das funktionierende Miteinander in einer Gruppe, die es so kein zweites Mal gibt. Vielen Dank für eure so vielseitige Unterstützung, liebe Ute Widow, Beate Jaschok-Kentner, Claudia Hanko, Angelika Arnscheidt und Simone Savio.

Zu guter Letzt möchte ich meiner Familie danken: Papa Helmut, Papi Stefan und Mama Margot, Andi, Clé, Juju, Sophia, Beate, Yvonne, Stefan & Jens mit allen Partnern und Kindern. Egal was kam, ihr wart immer da.

Diese Arbeit möchte ich meiner Familie widmen. Nicole und Sarah, ihr seid meine größte Freude und Ansporn.

11 Curriculum Vitae

PERSONAL INFORMATION

Name:	Maurice Günther Diwo
Date of birth:	13.10.1987
Place of birth:	Bad Homburg v. d. Höhe, Germany
Nationality:	German

PROFESSIONAL BACKGROUND

Since Jun 2020	**Technical Sales Specialist Europe**, Tosoh Bioscience GmbH

ACADEMIC BACKGROUND

Mai 2016 – Feb 2020	**PhD Student at Helmholtz Centre for Infection Research** “Towards structure elucidation of the Proline Reductase Complex from *Clostridioides difficile*”. TU Braunschweig, Germany
Jan 2015 – Mar 2015	**Erasmus+ Studentship at Biomedical Centre, Uppsala University** Uppsala Universitet, Sweden
Oct 2013 – Sep 2016	**Master of Science: Chemical Biology** „Strukturelle und funktionelle Charakterisierung von IκBNS/NFκB Komplexen“. TU Braunschweig, Germany. Grade: 1.4
Oct 2009 – Aug 2013	**Bachelor of Science: Biotechnology/Biopharmaceutical Technology** „Stabilitätsanalysen von Ba^{2+}-vernetzten Alginatkapseln unter simulierten physiologischen Bedingungen“. TH Mittelhessen, Germany. Grade: 1.9

MILITARY SERVICE

Dec 2007 - Sep 2009	**Voluntary military service** FGS Bayern (F217), 2. Fregattengeschwader, Wilhelmshaven

SCHOLAR BACKGROUND

Sep 2004 - Jun 2007	**High School Graduation (Abitur)** Saalburgschule Usingen – Berufliches Gymnasium, Grade: 2.4

SCIENTIFIC PUBLICATIONS

Maurice Diwo, Wiebke Michel, Philipp Aurass, Katja Kuhle-Keindorf, Jan Pippel, Joern Krausze, Christina Lang, Wulf Blankenfeldt, Antje Flieger. NAD(H)-mediated tetramerization controls the activity of *Legionella pneumophila* phospholipase PlaB. bioRxiv, (2020). doi: https://doi.org/10.1101/2020.09.01.246603.